

STIC-Biotech/ChemLib

58170

From: Walicka, Malgorzata  
Sent: Monday, January 14, 2002 10:53 AM  
To: STIC-Biotech/ChemLib

I would appreciate the search for:

SEQ ID NO: 1 and 2. Application No. 09/833,782, the earliest priority is 4/12/2000. Please use the pending and commercial data bases.  
Please RUSH!

Thank you,  
Malgorzata Walicka  
Art Unit 1652  
tel.305-7270  
Mali box 10C01  
Room 10D06

Point of Contact:  
Mona Smith  
Bio. Specialist  
GM1 12314 Tel: 800-3273

CRFE

STIC-Biotech/ChemLib

From: Chan, Christina  
Sent: Monday, January 14, 2002 11:53 AM  
To: STIC-Biotech/ChemLib; Walicka, Malgorzata  
Subject: FW:

**Please rush. Thanks Chris**

-----Original Message-----

From: Walicka, Malgorzata  
Sent: Monday, January 14, 2002 10:57 AM  
To: Chan, Christina  
Cc: STIC-Biotech/ChemLib  
Subject:

Christina, could you authorize this rush search?

Thank you,

Malgorzata

Searcher: M. Smith  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/4/02  
Date Completed: 11/5  
Searcher Prep/Review: 10  
Clerical: \_\_\_\_\_  
Online time: 10

TYPE OF SEARCH:

NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

---



**THIS PAGE BLANK (USPTO)**



/translation="MTLGRNPLQAMSSYVDGRNVLRLWDLSPQIKRRTEELIAQ  
TKQVYDGLMLIEEVTYENCIALADVEVKYIVERTMLDFPOHVSSDKREVAATEA  
DRLSRFDIEMSMREDIFURIVLKEETDLGKIKPEARYLEKSKMGRNGHLPPEQ  
VNEIKRPMKRRSELCIDFNKLNEDDTLVSFKAELGALPDDFIDLSLEKTDNKKYKI  
TLKYPHYFVPMKCCIPETRRKMAFNTRCKEENTIILOELLPLRAKVALKLGXSTH  
ADFLVEMNTAKSTHWTAPLDLDSOKLPLGEAREFRLNLKKKECEKGFVDDGKIN  
ANDLHYMTOTBELKYSVDQELKVEFLEVTEGLNLYQELLGLSPFOVTDARWN  
KSTLYTVKDKATGELVGFYLDLYPRECKYNHACFGLOPCLPLDPSRQMSVAALV  
VNESQPRAGRPSLLRDEVRTYFHEFGVHMOICATDFARFSGTNVDFRVEFVSOM  
LENVWVDDSLRRLSKHYKDGSPITDDLLKLVASLVNVTGLLTQLQIVLSVDSOLH  
TNTSDAASEYKYTEILGVAATPTGNPATFGLIAGGYDGYTYGLWSEVFSMDMF  
YSCFKKEGIMNEVGMKYNLLIKFGGSLDGMQLNQLKREPKNQKFLMSRLHAP"  
BASE COUNT 826 a 545 c 641 g 720 t  
ORIGIN

Query Match 87.9%; Score 1858.2; DB 4; Length 2732;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;  
QY 1 atgacgcccgggtgccttttggtgtgcgaagcctccgcagagttggtgtccaggatt 60  
DB 175 ATGATCGTCGGTGTCTTCGGTGGAGAGAAAGCAATGTCCTCTTCAGGCAATGTCCTCTAT 234  
QY 61 ttactcagaatcagcttagaagaagatgatgtctctctccttcaggcaatgtcttccat 120  
DB 235 TTATTGAGATGACTTTAGGAGAGAAAGCAATGTCCTCTTCAGGCAATGTCCTCTAT 294  
QY 121 actgtggtggaagaatgttttaagatggatctttaccagagcctccgcagagttggtgtccaggatt 180  
DB 295 ACTGTGGATGGCAGAAATGTTTAAAGTGGGATCTTTCACAGAGCAATTAAGAAAGA 354  
QY 181 actgaggagctcattgtgcagaccacacaggtgtacgtgtgtgtggaatgctcgtgatt 240  
DB 355 ACTGAGGAGCTGATTCGCGAGACCAACAGGTATGATGATATGGAATGCTTGACAT 414  
QY 241 gaggaagtaactacagagacgtctgcaggcactggcagatgtagaagttaagtata 300  
DB 415 GAAGAGGTAACTTATGAGAACTGTTTGCAGGCACTGGCAGATGTGGAAGTGAAGTATA 474  
QY 301 gtggaagagaccatgtagacttcccagcagatgtatctctccttcgacaaagaagtagagca 360  
DB 475 GTGGAACAGACCATGCTAGACTTCCCAAGCAGTGTCTCTGACAAAGATACAGACA 534  
QY 361 gcaagtacagaagcagacaaaagactttctgttttgatattgagatgagcatgagagga 420  
DB 535 GCAAGTACAGAAGCAGACAAAAGACTTTCTCGTTTGTATTTGATGAGATGAGCATGAGAGA 594  
QY 421 gatatttgagagaattgttcattacaggaacacgtgtatctctggggaagataaacct 480  
DB 595 GATATCTTTCTGAGAAATGTTCTGTTAAAGAGAGACTGTGATTTGGGGAAGATCAAGCCT 654  
QY 481 gagggcagacgatacttgaaagaatcaattaaatgggaaagaaaatgggtccatctt 540  
DB 655 GAGCCAGAGAGATCTTGAAGAAAACAGTTAAATGGGAAAAGAAATGGGCTCCATCTT 714  
QY 541 cctgacaagtacagaatgaaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600  
DB 715 CGGGAACAAGTACAGAAATCAAGCAATGAGAAAGAAATGAGTGAGGTATGTAT 774  
QY 601 gattttaaaaaaacctcaatgagatgataccttcttctgtattttccaaggctgaact 660  
DB 775 GACTTTCAACAAAACCTCAATGAGGACGATACCTTCTTGTATTTTCCAAAGGCTGAACCT 834  
QY 661 ggtgctctcctgatttccattgacagtttagaagaagacagatgagcaagataaaa 720  
DB 835 GGTGCTCTCTGATGATTTCAATGACAGTTTTAGAAAAGACAGATGACAAAGTATATA 894  
QY 721 attaccttaaatatccacactatttccctgtcatgaagaagtgtgtatccctgaagacc 780  
DB 895 ATTACCTTAAATAATCCACACTATTTTCTGTCATGAGAAATGTTGTATTCCTTGAAACC 954  
QY 781 agaagaagatggaaatggcttttaatacaagggtgcgaagggaacaccacataattttg 840

DB 955 AGAAGGAAAATGAAAATGGCTTTTAATACAGGTGCAAGAGGAAAACACCAATACTCTG 1014  
QY 841 cagcactactccactcgcgaacaaagtggtgcaaaactactcgtgttatagcacacatgct 900  
DB 1015 CAGAGCTACTCCACTCGGGCAAGTAGCCAAACTACTTGGGCTACAGCACACATGCT 1074  
QY 901 gacttgccttgaaatgaacactgcgaagagcacaagcgcgcgttaacagccttcttagat 960  
DB 1075 GACTTTGCTCTTGAGATGAACACTGCGAAGAGTACACACCATGTGACAGCCTTCTAGAT 1134  
QY 961 gatttaagcagaagttaaaaccccttggtggaagcagaagaggtttattttgaatttg 1020  
DB 1135 GATTTAAGCCAGAAATTAATAACCTTGGGTGAGCGCAAGACGAGAGTTCAATTTGAATTTG 1194  
QY 1021 aagaaaaagaaatgcaagacagaggggttttgaatatgatgggaaaaataatgactcgtggat 1080  
DB 1195 AAGAAAAGGAATGTGAAGAGAAAGGTTTGAATATGATGGCAAAATCAATGCTGGGAT 1254  
QY 1081 ctatatctactactgactcagacagaggaactcaagtatttccatagaccaagagttccctc 1140  
DB 1255 CTACATTACTATGACTCAGACAGAGAACTCAAGTACTCTGTGGACCAAGAGATGCTTC 1314  
QY 1141 aaggaatacttcccaattgaggtggtcactgaagcgttgcacacacctaccagaggttg 1200  
DB 1315 AAGGAATACTTCCCAATTAAGCTGCTCACTGAAGGCTTCTGCAACATCTATCAGGAGTTG 1374  
QY 1201 ttgggaactttcattgaacaaaatgacagatgctcattgtttggaacaaagagtttaccatt 1260  
DB 1375 TTGGGACTCTCCTTTGAGCAAGTACAGATGCTCATGTTTGGAAATAAAGTGTTCACCTG 1434  
QY 1261 tatactgtgaagataaagctacagagaagatattgggacagttcttatttgacctctat 1320  
DB 1435 TACACGGTGAAGGATTAAGCCCAAGAGAGAAATATTAGGACAGTTTCTACTTGGACCTCTAT 1494  
QY 1321 caaaggaaggaataacatacatcgccgtcctcgtgctcctcagcgtggtgcctctcg 1380  
DB 1495 CCAAGGAAGGAAAATACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1554  
QY 1381 cctgatggaagccgagatgatggcagtggtgctgctcctcgttggaacttctcacagccagt 1440  
DB 1555 CCGTATGGAGCGCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1614  
QY 1441 gcagctcgtccctcctcctcagacacagcagagtgagagacttaacttcaatgagtttgt 1500  
DB 1615 GCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1674  
QY 1501 cagctgtagcatcagattgtgcacagactgattttgcaacagcttagcggaacaaatgtg 1560  
DB 1675 CACGTCATGCATCAGATTTGTGGCGAGACTGACTTTTGGCAGATTTAGTGGAAACAATGTG 1734  
QY 1561 gaaactgactttgtagagtgccatcgcaaatgcttgaaaattgggtgtggagactcgtat 1620  
DB 1735 GAAACTGACTTTGTAGAGTGCATCACAAAATGCTGTGAAAACCTGGGTGTGGGACACTGAT 1794  
QY 1621 tccctccgaagatgtcaaaaacattataaagatggaagccctattgagagatcgtctt 1680  
DB 1795 TCCCTCCCAAGGCTGTCAAAACATTTATAAGATGGAAGCCCTATTACAGATGATGCTGCT 1854  
QY 1681 gaaaaactgttgcctcgtggtgtaacacagagctctctgacccctgcgcagatgatt 1740  
DB 1855 GAAAAACTTGTGCTCTAGACTGGTCAACACAGAGTCTCCGACCTTACGCCAGATGCTT 1914  
QY 1741 ttgagcaagtgtgactcgtctcttcatacccaacacatcgctggatgctgagtgaaatgaa 1800  
DB 1915 TTGAGCAAGTTGATGACTCTCCACACCAACTTCTCGCTGGATGCTGCAAGTGAATAT 1974  
QY 1801 gccaaatctcctcagaaaattattaggaggttcagctactccagccacaaatattgccagct 1860  
DB 1975 GCCAAATACTGCACAGAAATTTTAGGTGTTGACAGTACTTCCAGGCACAAAATATGCCAGCT 2034  
QY 1861 accttggacatttggcaggggagatagatggcgaatattatgatcttcttggagtaa 1920

**THIS PAGE BLANK (USPTO)**

```

Db 2035 ACTTTGGACACTGGCAGGGGATATGATGGCCAATATTATTGATATCTTTGGAGTAA 2094
QY 1921 gtatttccatgatgtttttacagctgttttaaaaaaagaggataatgaatccagag 1980
Db 2095 GTCTTTCCATGACATGTTTATAGCTGTTTAAATAAAGAGGATAATGAATCCAGAG 2154
QY 1981 gttggaaatgaataacacagaaacctaactcgaacacctgggggagctctctggagcgcatggag 2040
Db 2155 GTTGAATGAATACAGAAACCTAATCTCAAAACCTGGGGGATCGCTGGAGCGCATGGAC 2214
QY 2041 atgctccacaatttcttgaacagtgagcacaacacaaacagcttctcctaagtagagggc 2100
Db 2215 ATGCTCCAAAATTTCTTGAACGTGAGCCAAACCAAAACCGTTCTCTTAATGAGTAGAGGC 2274
QY 2101 ctgcatgctcgtgaa 2115
Db 2275 CTGCATGCTCCGTA 2289

RESIDUES 4
LOCUS PIGSABP 3819 bp mRNA MAM 14-APR-2000
DEFINITION Sus scrofa mRNA for soluble angiotensin-binding protein, complete cds.
ACCESSION D11336
VERSION D11336.1 GI:217708
KEYWORDS SINE; sABP; short interspersed repeated sequence; soluble angiotensin-binding protein.
SOURCE Sus scrofa (sub-species:domesticus) heart, liver cDNA to mRNA, clone_11b:lambdazap II clone:lambdapAB[1.4,s2,-L1].
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 3819)
AUTHORS Sugiura,N., Hagiwara,H. and Hirose,S.
TITLE Molecular cloning of porcine soluble angiotensin-binding protein
JOURNAL J. Biol. Chem. 267 (25), 18067-18072 (1992)
MEDLINE 9238174
REFERENCE 2 (bases 1 to 3819)
AUTHORS Hirose,S.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-1992) to the DDBJ/EMBL/GenBank databases.
Shigehisa Hirose, Tokyo Institute of Technology, Department of Biological Sciences, Ookayama, Meguro-ku, Tokyo 152, Japan (Tel:03-3726-1111(ex.2226), Fax:03-3729-0335)
FEATURES
source
1..3819
/organism="Sus scrofa"
/sub_species="domesticus"
/db_xref="taxon:9823"
/clone="lambdapAB[1.4,s2,-L1]"
/clone_11b="lambdazap II"
/tissue_type="heart, liver"
158..2272
/note="sABP"
/codon_start=1
/product="soluble angiotensin-binding protein"
/protein_id="BA01949.1"
/db_xref="GI:217708"
CDS
/translation="MIVRCLSAARLHRVGGSGILLRMTLIGREAMSPLOAMSYTVDG
RNYLRWLSPEQIKRRTEDLQKQVDDIGMLDEEYVENCLOALADVEVKYIVE
RTMLDPPOHVSQKVEYRAAETADKRLSRFDIEMSMREDIFLRIVRHETCDLQKTP
EARYLEKSVKMGKRNGLHPEQVNEIKAMKRMSELCDENKLNEDDTFLVFSKA
ELGALPDFIDSLKEDDNKIKITLPIFYFVNNKCCIPETRRKMEMAFNTRCKEEN
TIILQELLURAKVAKLGYSHADFLVEMNTAKSHVYFADLDDLSOKLPIEGEER
EFTLNKKECEKEFYDGIKINADLHYMTQTIELKYSVDQIELKYPPIEVFTEG
LLNIYDLGLISFEQVTDVHWNVKSLTVKDKATGEVLQGYLDLYPREGYNHAA
CFGLQPCLLPDGSRMVAALVYVNSQPRAPRSLLRHDEVTYFHEGHNHQA
QTFAPFSGNVEFDVFPVSQMLENWDTSRLRSKHKYGCSPTDLDLKLVAS
RLVNTGLLURQILSKVDQSLHNTSLDASEYATCYTEILGVAATPGTNMPATFGH
LAGGYDQYQYGLWSEYFSDMFYSCFKREGIMNPEVGMKRYRLILKPGGSLDGMDL
QNFLKREPNGKQFLMSRGLHAP"
2001..2137
variation

```

```

variation
/note="deleted from clone lambdapAB1"
2168..2169
/note="substituted in clone lambdapAB1"
2676..2681
polyA_signal 2989..2994
polyA_signal 3459..3780
repeat_unit
/note="SINE(PRE-1)"
BASE COUNT 1086 a 788 c 855 g 1090 t
ORIGIN
Query Match 87.9%; Score 1858.2; DB 4; Length 3819;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;
QY 1 atgatacgcgcgagtccttttgctgtgcgaagccctccgcagagttggtgttccaggatt 60
Db 158 ATGATCGTCCGTGCGCTTTCGGCTCGGAGAGCCCTCCACAGGCTGGTGTCTCCGGATT 217
QY 61 ttactcagaatgacgttaggaagagaagtgtgtctctcttcaggcaaatgtcttctat 120
Db 218 TTATTGAGATGACITTAGGGAGAGAGCAATGTCTCTCTTCAGSCAATGTCTTCTAT 277
QY 121 actgtgctggcagaaatgttttaagatgggattctttccaccagagcaaatataaacaaga 180
Db 278 ACTGTGGATGGCAGAAATGTTTAAAGATGGGATCTTTCACCAGAGCAAAATTAAGAAGA 337
QY 181 actgagagactcatgtgcagaccacaaagggtgacgatgctgttggaatgctcgatt 240
Db 338 ACTGAGGAGCTGATGCGCAGACCAACAGGTGTATGATGATTTGGATGCTTGACATT 397
QY 241 gaggaagtaacttcacgagaactgtctgcaggcactgcgcagatgtagaagtaaatataata 300
Db 398 GAAGAGGTAACTTATCAGAACTGTTTTCAGGCACCTGCGAGATGTGGAAGTATATA 457
QY 301 gtggaaggaacatgttagactttcccccagcatgtatctctctgacaaaagaagtacagca 360
Db 458 GTGGAACGAACCATGCTAGACTTTCCGCCAGCATGTCTCTCTGACAAAGAAGTACGAGCA 517
QY 361 gcaagtacagaagcagacaaaagactttctgtttgattgagatgagcatgagcagagga 420
Db 518 GCAAGTACAGAACGACACAAAGACATTTCTGTTTGTATTTGATGATGAGATGAGAGAA 577
QY 421 gatattttgagaattgttctatttacaggaaacctgtgatctggggagagataaaacct 480
Db 578 GATATCTTTTGAGAAATGCTTCGTTTAAAGGACACCTGTGATTTGGGGAAGATCAAGCCT 637
QY 481 gagggcagacgatacttggaaaaagtcaattataaatgggaaaaagaatgggctccatctt 540
Db 638 GAAGCCAGAAGATACTTAGAAAAATCAGTTAAATGGAAAAAGGAATGGGCTCCACTT 697
QY 541 cctgaacaagtacagaatgaatcaaatcaatgaagaaaaagaatgagtgagctatgatt 600
Db 698 CCGGAACAAGTACAGAAATGAATCAAAAGCAATGAAGAAAAAGATGAGCTGATGATT 757
QY 601 gattttaacaaaaaacctcaatgaggatgataccttctgtatttttccaaaggtgaactt 660
Db 758 GACTTCAACAAAAACCTCAATGAGGACGATACCTTCTCTGTTATTTCCAGAGGTGACTT 817
QY 661 ggtgctcttctgatgatttcttcagcagtttagaagaacacagatgatgacaagtataaa 720
Db 818 GGTGCTCTCTCTGATGATTTCAATGACAGTTTAAAGAACACATGACACAACAGTATAAA 877
QY 721 attacctataaatatcacacactatttccctgcatgaagaatgtgtatccctgaacc 780
Db 878 ATTACCTTAAATATCACACTATTTTCTGTCTCATGAAGAAATGTTGTATCCCTGAAACC 937
QY 781 agaagaagatgaagaatgcttttaatacaagtgcaagagaaaaaacacataatttg 840
Db 938 AGAAGGAAATGGAATGCGCTTTTAAACAGGTGCAAGAGGAAACACCATAAATCCTG 997
QY 841 cagcagctactcccactgcgaaccaaaggtggccaaactactcgtgtatagcacacatgct 900

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 14, 2002, 18:59:51 ; Search time 3021.82 Seconds  
(without alignments)  
11546.530 Million cell updates/sec

Title: US-09-833-782-1  
Perfect score: 2115  
Sequence: 1 atgatgccgcgggctctttt.....gagccctgcactcgtcgtga 2115

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_fun:\*
- 18: em\_in:\*
- 19: em\_om:\*
- 20: em\_or:\*
- 21: em\_ov:\*
- 22: em\_pat:\*
- 23: em\_ph:\*
- 24: em\_pl:\*
- 25: em\_ro:\*
- 26: em\_sts:\*
- 27: em\_sy:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htgo\_hum:\*
- 31: em\_htgo\_inv:\*
- 32: em\_htgo\_rod:\*
- 33: em\_htg\_hum:\*
- 34: em\_htg\_inv:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	2114.2	100.0	2890	9	HS300837	AJ300837 Homo sapi
2	2109.4	99.7	2864	6	AX106913	AX106913 Sequence
3	1858.2	87.9	2732	4	AB000170	AB000170 Porcine m
4	1858.2	87.9	3819	4	PIGSABP	D11336 Sus scrofa
5	1829.8	86.5	2810	4	AB000172	AB000172 Porcine m
6	1828.4	86.4	2709	4	AB000171	AB000171 Porcine m
7	1828.4	86.4	2862	4	AB000173	AB000173 Porcine m
8	1828.2	86.4	2604	4	AB000174	AB000174 Porcine m
9	1826.8	86.4	2728	4	AB000175	AB000175 Porcine m
10	1799	85.1	3507	4	RABEPD	D13310 Rabbit mrna
11	1653.4	78.2	2448	10	RNRNAME	X87157 R.norvegicu
12	1526	72.2	2637	9	BC004985	BC004985 Homo sapi
13	1526	72.2	2646	9	BC001644	BC001644 Homo sapi
14	1344.2	63.6	5343	9	AB03052	AB03052 Homo sapi
15	1178.6	55.7	1658	9	AK027602	AK027602 Homo sapi
16	857.6	40.5	2850	5	AB030904	AB030904 Xenopus l
17	800.4	37.8	2490	4	PIGTHOL	D21871 Pig mrna fo
18	790.8	37.4	2070	6	I12383	I12383 Sequence 6
19	790.8	37.4	2462	9	HSU29366	U29366 Human Chine
20	790.8	37.4	2515	9	HSTH10PS	250115 H.sapiens m
21	790.8	37.4	2535	9	BC000135	BC000135 Homo sapi
22	790.8	37.4	2577	9	BC002391	BC002391 Homo sapi
23	779.6	36.9	2419	10	AF314187	AF314187 Mus muscu
24	759.2	35.9	2314	10	RATMEP	M61142 Rat metallo
25	694.6	32.8	2122	9	BC000583	BC000583 Homo sapi
26	372.8	17.6	138295	2	AC024448	AC024448 Homo sapi
27	372.8	17.6	167656	9	AC079468	AC079468 Homo sapi
28	372.8	17.6	179969	9	AC016643	AC016643 Homo sapi
29	372.8	17.6	196898	2	AC010357	AC010357 Homo sapi
30	321	15.2	391	4	AB00041S10	AB000420 Porcine D
31	263.8	12.5	109525	9	AC008958	AC008958 Homo sapi
32	263.8	12.5	196898	2	AC010357	AC010357 Homo sapi
33	226.4	10.7	284	4	AB00041S04	AB000414 Porcine D
34	180.4	8.5	138295	2	AC024448	AC024448 Homo sapi
35	179.4	8.5	226	4	AB00041S11	AB000421 Porcine D
36	172.2	8.1	204	4	AB00041S12	AB000422 Porcine D
37	153.6	7.3	3815	8	SCPRD1	X76504 S.cerevisia
38	153.6	7.3	316613	8	SCCHRII	X59720 S.cerevisia
39	152.2	7.2	185	4	AB00041S08	AB000418 Porcine D
40	149.6	7.1	385	4	AB00042S08	AB000433 Porcine D
41	139.4	6.6	173	4	AB00041S05	AB000415 Porcine D
42	133.2	6.3	112022	9	HSJA9611	AJ009611 Homo sapi
43	133.2	6.3	177540	9	AC006538	AC006538 Homo sapi
44	132.4	6.3	201320	2	AC073816	AC073816 Mus muscu
45	132.4	6.3	247196	2	AC073822	AC073822 Mus muscu

ALIGNMENTS

RESULT 1  
HSA300837  
LOCUS HSA300837 2890 bp mrna  
DEFINITION Homo sapiens mRNA for neurolysin.  
ACCESSION AJ300837  
VERSION AJ300837.1 GI:12583611  
KEYWORDS neurolysin.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 2890)  
AUTHORS Chen, J. M., Rawlings, N. D. and Barrett, A. J.  
TITLE Cloning and sequencing of human neurolysin, an oligopeptidase of family M3  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2890)  
AUTHORS Chen, J. M.  
TITLE Direct Submission  
JOURNAL Submitted (23-JAN-2001)-Ghent-U.M.-MRC-Molecular-Enzymology

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

Laboratory, The Babraham Institute, Babraham, Cambridge, CB2 4AT,  
UNITED KINGDOM

FEATURES  
source  
1. .2890  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
126. .2240  
/EC\_number="3.4.24.16"  
/function="oligopeptidase"  
/codon\_start=1  
/evidence="experimental"  
/product="neurolysin"  
/protein\_id="CAC27329.1"  
/db\_xref="GI:12583612"  
/translation="MIARCLLVSLRRVSGSRLLRLMTLGRVMSPLQAMSSVTVAG  
RNVLRWLSPEQIKTRTEELIVQTKYQVDAVGLGIEEYVENCLOALADVKEYIIVE  
RTMLDFQHVSSDEKRAASTEDKRLSRDIEDMSMRGDIIFERIVHQLQEDVLKIKP  
EARYLEKSRMGKRLHLEPEQVNEIKSMKRMSLCLDFNKNLNMEDDTFLVFSKA  
ELGALPDFIDSLKDDDKYKITLKYPHYFPVMKKCCIPETRRMEMAFNTRKEEN  
TIILOQLPLRTKVKALGLYSTHADVLENNATKSTKSTAFDLDDLKSLKPLGEAR  
EFILNKKKECKRFRFEDGKINAWDLIYYMTOTEEKYSIDOEFLKEYPEIEVVTG  
LLNTYQELGLSPQMTDAHWNNKSVLYFVKDKATGEVLGQFLDLIYPREGYNHAA  
CFGLQPCLLPDGSRMAVAALVYVNSQPVAGRPSLRKHDEVRTYFHEFGVHMQICA  
QTFARFSGNVEFDEVEFVPSQMLENNWVDVSLRSLKSKYKDGSPHADLLEKLVAS  
RLVNTGLTLQIVLSKVDQSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGH  
LAGGYDQGYLYLWSEVFSMDMFYSCFKKEGIMNPEVMKYRNLLIKPGGSLDGMML  
HNFLKRPNQKAFMLSRGLHAP"

BASE COUNT 861 a 571 c 562 g 796 t

ORIGIN

Query Match 100.0%; Score 2114.2; DB 9; Length 2890;  
Best Local Similarity 99.98; Pred. No. 0;  
Matches 2113; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgacgcccgtgctcttggctgtgcgaagcctccgacagtggtgggtccaggatt 60  
Db 126 ATGATCGCCGCTGCTTTTGGCTGTGCGAAGCCTCCGACAGTGTGGTTCAGGATT 185

Qy 61 ttactcagaatgaagtaggaagaagtagtctctctctctccttcaggcaattcttcctat 120  
Db 186 TTACTCAGAATGACGTTTAGGAAGAAGTATGCTCTCTCTTCAGGCAATGCTCTTCCTAT 245

Qy 121 actgtggctgcagaaatgttttaagtaggattcttccacagagcaaaataaaacaaga 180  
Db 246 ACTGTGGCTGCAGAAATGTTTTAGATGGGATCTTTCACCAGAGCAAAATTAACAACAGA 305

Qy 181 actgaggagctcattgtgcagacaaacaggtgtacagatgctgtgtggaatgctcggtatt 240  
Db 306 ACTGAGGAGCTCAITGTGCAGACCAACACAGGTGTACGATGCTGTGTGAATGCTCGGTATT 365

Qy 241 gagaaagtaactacagaaactgtctgcaggaactgtgcagatgtagaagtaaaagtatata 300  
Db 366 GAGAAAGTAACTTACGAGAACTGTCTGCAGCACTGGCAGATGTAGAAGTAAAGTATATA 425

Qy 301 gtgaaagaccatctagacttcccacagatgtatctctctcgaagaagaagtagcagca 360  
Db 426 GTGAAAGACCATCTAGACTTTTCCCAGCATGTATCTCTGACAAAGAAGTACGAGCA 485

Qy 361 gcaagtcagaaagcagacaaagaacttctctgttttgatatgtgagatgagcatgagagga 420  
Db 486 GCAAGTCAGAAAGCAGACAAAGACTTCTCGTTTTGTATTTGAGATGAGCATGAGAGGA 545

Qy 421 gatattattgagaattgttcatcttacagaaacccctgtgtctctgggaaagataaaacct 480  
Db 546 GATATATTGAGAATTTGTTTCAATTAACAGAAACCTGTGTATCTGGGGAAGATAAAACCT 605

Qy 481 gagccagacgatacttggaaagtcaattaaatgggaaagaaataatgggctccattctt 540  
Db 606 GAGCCAGACGATACTTGGAAAGATCAATTAATATGGGAAAGAAATATGGGCTCCATCTT 665

Qy 541 cctgaacaggtacagaatgaatacaatcaatgaagaaagaatgagtgagctatgtatt 600

Db 666 CCTGAACAAGTACAGAATGAATCAATCAATGAAGAAAGAAATGAGTGAAGTATGATT 725  
Qy 601 gatTTtaaaaaaacctcaatgagatgatcaccttcccttctgtatttccaaagcgtgaacct 660  
Db 726 GATTTTAAACAAAACCTCAATGAGGATGATACCTCTCTGTATTTCACAGCGTGAACCT 785  
Qy 661 ggtgctctcctgatgattcattgacagtttagaagaagacagatgatgacaaagtataaa 720  
Db 786 GGTGCTCTTCCTGATGATTTCATTGACAGTTTAGAAAAGACAGATGATGACAAAGTATAAA 845  
Qy 721 attaccttaaaatatccacactatttccctgtcatgaagaagaattgttgcctcgaacc 780  
Db 846 ATTACCTTTAAATATCCACACTATTTCCTGTCATGAAGAAATGTTGTATCCCTGAAACC 905  
Qy 781 agagaagagtgaaaatggctttlaatacagaagtgcaagaaggaagaaacaccataatttg 840  
Db 906 AGAAGAAGATGGAATGGCTTTTAAACAAGGTGCAAGAGGAAACACCATTAATTTTG 965  
Qy 841 cagagctactcccactgcgaaccaaggtggccaaactactcgggttagcacacatgct 900  
Db 966 CAGCAGCTACTCCACACTGCGAACAAGGTGCGCAAACTACTCGGTTATAGCACACATGCT 1025  
Qy 901 gacttcgtccttgaatgaacactgcgaagcgcacaaagcgcgcgttaacagcgtttctatag 960  
Db 1026 GACTTCGTCCTTGAATGAACACTGCAAGAGAGCACAAAGCCGCTAACAGCCTTTCATAG 1085  
Qy 961 gatttaagcagaagttaaaaacccttgggtgaagcagaacagagatttatttgaatttg 1020  
Db 1086 GATTTAAGCCAGAGTTTAAACCCCTTGGGTGAAGCAGAACGAGAGTTTATTGTAATTG 1145  
Qy 1021 aagaaaaaggaatgcaaaagacaggggttttgaatatgatgggaaatacaatgcctgggat 1080  
Db 1146 AAGAAAAGGAATCCAAAGACAGGGGTTTTGAATATGATGGGAAATCAATGCTGGGAT 1205  
Qy 1081 ctatattactacatgactcagacagagaaactcaagtattccatagacaaagagttctctc 1140  
Db 1206 CTATATTACTACATGACTCAGACAGAGAACTCAAGTATTCCATAGACCAAGAGTTTCCTC 1265  
Qy 1141 aagaaatacttcccactgaagtggtcactgaagcttgcgtgaacacaccacacagagttg 1200  
Db 1266 AAGAAATACTTCCCAATGAGTGGTCACTGAAGCTTGTGTAACACCTCACAGAGATTG 1325  
Qy 1201 ttggagacttcttgaacaaatgacagatgctcatgtttggaaacagagttgtacacct 1260  
Db 1326 TTGGGACTTTTCATTGAAACAAATGACAGATGCTCATGTTTGAACAGAGGTTTACACTT 1385  
Qy 1261 tatactgtaaggataaagctacaggaagaatattgggaacagttctatttggaaacctctat 1320  
Db 1386 TATACTGTGAAGGATAAAGCTACAGAGAGAGTATTGGACAGATTCTATTGAGCCTCTAT 1445  
Qy 1321 ccaaggaagaagaataacaatcatgcggcctgcttcgggtctccagcctggctgctctctg 1380  
Db 1446 CCAAGGGAAGGAAATAACAATCATCGGCGCTGCTTCGGTCTCCAGCCCTGGCTTCTG 1505  
Qy 1381 cctgatgaagccggatgatggcagtgctgctcctcgtgtggaactctcacaagcagtg 1440  
Db 1506 CCTGATGGAAGCCCGGATGATGGCAGTGGCTGCCCTCGTGGTGAACCTCTCACAGCCAGTG 1565  
Qy 1441 gcaggtcgtcctctcctctgagacagcagagtgaggacttactttcatagatttgggt 1500  
Db 1566 GCAGGTGCTCCTCTCTCCTGAGACACAGCAGGAGTGAGGACTTACTTTTCATGAGTTTGGT 1625  
Qy 1501 cactgtatgatacagatttgcacagactgatttgcacagatttgcagaaacaaatgtg 1560  
Db 1626 CACGTGATGATCAGATTGTGACAGACTGATTTTGCACGATTTAGCGGAACAAATGTTG 1685  
Qy 1561 gaaactgactttgtagaggtgccaatgcataatgcttgaataattgggtgtggagcgtcgat 1620  
Db 1686 GAAACTGACTTTGTAGAGGTGCCATGCAAAATGCTTGAATAATTGGTGTGGGACGTCGAT 1745  
Qy 1621 tccctccgaagattgtcaaaacataataaagatggaagccctattgcagacgactctgctt 1680

Db 1746 TCCTCCGAAGATTGCAAAACATTTATAAGATGAAGCCCTATTTCAGACCATCTGCTT 1805

Qy 1681 gaaaaactgttcttcttaggtgtcaacacaggtcttcttgcacctgcgcagattgtt 1740

Db 1806 GAAAAACTTGTCTTCTAGGCTGTCACACAGCTCTTCTGACCTGCGCCAGATTGTT 1865

Qy 1741 ttgagcaaatgtgatcagctcttctataccacacacatgcgtgatgctgcaagtgaat 1800

Db 1866 TTGAGCAAAAGTTGATCAGTCTCTTATACCAACACATCGCTGGATGCTCAAGTGAATAT 1925

Qy 1801 gccaaatactgtcagaaataattaggttgagctactctccagggcacaaatgcccagct 1860

Db 1926 GCCAATACTGTCAGAAATATTAGAGTTCAGCTACTCCAGGCACAAATATGCCAGCT 1985

Qy 1861 accttggacatttggcaggggtatcagtggtgccaataattatggatatcttggagtga 1920

Db 1986 ACCTTTGGACATTTGGCAGGGGATACGATGTCGCAATATTATGATATCTTTGGAGTGA 2045

Qy 1921 gtaattccatgatattttacagctgttttaaaagagaggaatgaatgaatgaatgaat 1980

Db 2046 GTATTTTCCATGATGATGTTTACAGCTGTTTAAAGAGAGGGATATGATGATATCTTT 2105

Qy 1981 gttggaatgaatlacagaacctaactctgaaacctggggggtctctgacgcatggac 2040

Db 2106 GTTGGATGAATACAGAACTAATCTTGAACCTGGGGGATCTCTGACGCGCATGGAC 2165

Qy 2041 atgtccacaattcttgaacgttgagccaaacccaaagcgttcttaattagtagagac 2100

Db 2166 ATGCTCCACAATTTCTGAACGTGAGCCAAACCAAAAGCCGTTCTTAATGATGATAGGC 2225

Qy 2101 ctgcatgctctgtga 2115

Db 2226 CTGCATGCTCCGGA 2240

RESULT 2

LOCUS AX106913 2864 bp DNA PAT 30-APR-2001

DEFINITION Sequence 2 from Patent WO0123590.

ACCESSION AX106913

VERSION AX106913.1 GI:13922466

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Kapeller-Libermann, R., White, D. and Silos-Santiago, I.

TITLE 22196, a novel human aminopeptidase

JOURNAL Patent: WO 0123590-A 2 05-APR-2001

FEATURES

source Location/Qualifiers

1..2864

/organism="Homo sapiens"

/db\_xref="taxon:9606"

300..2414

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC37750.1"

/db\_xref="GI:13922467"

/translation="MIARCLLAIVSLRRVGGSRILLRMTLGRVMSPLQAMSSYTVAG RNVRDLSPQIKTTELEIVQTKQYDAVGLMGEIVTYENCLOALADVEKVIYE RTMLDPQIVSSDKRNLRASTADKRLSRFDEMSMRGDI FERI VHLQETDLGKIKP EARYLEKIRKQKRNGLHPQVQNEIKSMKRMSELIDFNKNLNEDDTTLVFSKA ELGALPDDTDLSEKTDKDYKTLKYPHYFVPMKCCIPETRRRMEAFNTRCKREN TILQOLLPLRTKVLKLLGYSTHADVLEMTAK STSRVTAFLDDLSOKLPLGEAER EFLNLKKECKDRGEYDGIKINAWDLYYMTQTELEKYSIDQELKEYPLEVTYEG LNTYQELGLLSEFQMDAHVWNSVLTIVTKDKATGEVIGQFYLLDYPREGKYNHAA CFTGLQGLLSPGRMMVAALVNFSPVAGRSLLLRHDEVRTYFPHFGVHMQICA QTFDFAFGSTNVETDFEVSQMLENWNVDVSLRSLKSKYDGS IADDLLEKIVAS RLAVNTGLLTROIIVLSKVQSLHTSLDAASEYAKYCSILLGVAATPGTNMPATFGH LAGYDQGVYGVYLSVFSVMDMFYSCFKKEGIMNPEVGMKRYNLILKPGSLDGMDL HNFLKREPNOKAFILMSRGLHAP"

CDS

BASE COUNT 854 a 556 c 693 g 761 t

ORIGIN

Query Match 99.7%; Score 2109.4; DB 6; Length 2864;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2110; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgctgccccgggtgctcttggctgtgcgaagcctccgcagagttgggttccaggatt 60

Db 300 ATGATGCGCCGGTGGCTTTGGCTGTGCGAAGCCCTCCGAGAGTTGGTGTTCAGGATT 359

Qy 61 ttactcagaatcagcttaggaagaagtgatgctcctcttcaggcgaatgtcttctctat 120

Db 360 TTACTCAGNATCAGCTTAGGAAGAAGATGATGCTCTCTTCAGGCATGCTTCCCTAT 419

Qy 121 acttggctggcagaatgttttaagatgggatcttccaccagagcaaatataaacaaga 180

Db 420 ACCGTGGCTGGCAGAAATGTTTTAAGATGGGATCTTTCCCGCAGAGCAAAATTAACAAGA 479

Qy 181 actgaggagctcattgtgcagaccacaaacaggtgtagctgctgttggaaatgctcgatt 240

Db 480 ACTGAGGAGCTCATTGTGCAGACCAGAGGTGTAGATGCTGTGGAAATGCTCGGATT 539

Qy 241 gaggaagtaacttaagagaactgtctgcaggcactgcgcagatgtagaagtaaatata 300

Db 540 GAGGAAGTAACCTTACGAGAACTGTCTGCAGGCACCTGGCAGATGTAGAACTAAGTATA 599

Qy 301 gtggaaaggaccatgctagacttcccagcatgtatcctctgacaaagaagtaagca 360

Db 600 GTGAAAGGACCATGCTAGACTTTCCCGCAGCATGTATCTCTGACAAAGAGTACGAGCA 659

Qy 361 gcaagtacagaagcagacaaaagacttctctgttttgatattgagatgagcatgaagca 420

Db 660 GCAAGTACAGAAGACAGCAAAAGACTTTCTCGTTTGTATTGAGATGAGCATGAGAGGA 719

Qy 421 gatatttgagagaattgttcattcacaggaaacctgtgactctgggggaagataaaacct 480

Db 720 GATATATTGTAGAGAATTGTTTCATTTACAGGAACCTGTGTCTGGGGAAGATAAACCT 779

Qy 481 gaggccagacgatacttggaaagtcaatataaatggggaaagaagaatgggtccatctt 540

Db 780 GAGGCCAGACGATACCTTGGAAAGTCAATTAATAATGGGAAAGAAATGGGTCCATCTT 839

Qy 541 cctgaacaagtcagacgaatgaatacaaatcaatgaagaagaagaatgagtgatctatt 600

Db 840 CCTGAACAAGTACAGAAATGAAATCAATCAATGAAGAAAGAAATGAGTGTGTATT 899

Qy 601 gattttaacaaaaacctcaatgagatgataccttctctgtattttccagggtgaactt 660

Db 900 GATTTTACAAAACCTCAATGAGGATGATACCTCTCTGTATTTCCTCAAGGCTGAACCT 959

Qy 661 ggtgctcttctgattgatttcttgcagtttagaagaagacagatgatacagaatataa 720

Db 960 GGTGCTCTCTCTGATGATTTTCATTTAGACAGTTTAGAAAAGACAGATGATGACAAAGTATA 1019

Qy 721 attacttataataatccacactatttccctgtctatgaagaagaatgttgcctgaacc 780

Db 1020 ATTACCTTAAATATCCACACTATTTCCCTGTCTATGAAGAAATGTTGTATCTCCCTGAAC 1079

Qy 781 agaagaagatgaaatggcttttaatacaaggtgcaagaaggaacacacataattttg 840

Db 1080 AGAAGAAGATGGAATGGCTTTTAATACAGGTGCAGAGGAGGAAACACCAATATTTG 1139

Qy 841 cagcagctactccactgcgaacaaagggtggccaaactactcgtgttatagcacacatgct 900

Db 1140 CAGCAGCTACTCCCACTGCGAACCAAGTGGCCAAACTACTCTCGTTTATAGCACACATGCT 1199

Qy 901 gacttgccttgaatgaacactgacaaagcagacagccgcgttaacagcgtttcttagat 960

Db 1200 GACTTCTCTCTGAAATGAACACTGTCAAGAGCAGCAAGCCGCTTAACAGCTTCTTAGAT 1259

Qy 961 gatttaagccagaagttaaaaccttgggtggaagcagaacgagattttttgaattttg 1020

Db 1260 GATTTAAGCAGAAAGTAAACCCCTTGGTGAAGCAGAGAGTATTTTGAATTG 1319  
QY 1021 aagaaagaaatgcaaacagaggttttgaatgatggaagaaatcaatcggtgat 1080  
Db 1320 AAAAAAAGAAATGCAAGACAGGGGTTTTGAATATATGATGGGAAATCAATGCGGTGAT 1379  
QY 1081 ctatattactacagactcagacagaggaactcaagtattccatagacacaaaggttccctc 1140  
Db 1380 CTATATTACTACATGACTCAGACAGAGGAACCAAGTATTCATAGACCAAGAGTTCTC 1439  
QY 1141 aagaaatacttcccaattgaggtggtcaatgaaggttctgtaaacacctaccaggagttg 1200  
Db 1440 AAGAAATACTTCCCAATTGAGGTGGTCACTGAAGGCTTCTGTAACACCTACCAGGAGTTG 1499  
QY 1201 ttggagacttcatgaaacaaatgacagatgctcatgtttgaaacaaagatttacaatt 1260  
Db 1500 TTGGGACTTTCATTGAACAAATGACAGATGCTCATGTTGGAAACAGAGTGTACATT 1559  
QY 1261 tatactgtgaagataaaactacagagaggaattggtgacagttcttattggacctctat 1320  
Db 1560 TATACGTGTAAGGATAAAGCTACAGGAGAAGTATTTGGGACAGTTCTATTGGACCTCTAT 1619  
QY 1321 cbaaggaaggaataacataatcatcggtcgttgggttctccagcctggtgcttctg 1380  
Db 1620 CCAAGGGAAGGAAATPACAATCATGCGGCTGCTTGGGTCTCCAGCTGCGCTGCTTCTG 1679  
QY 1381 cctgatgaagcagatgatgcagtgagcagtggtggtggtggtggtggtggtggtggtggt 1440  
Db 1680 CTTGATGGAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739  
QY 1441 gcaagtgctgctctctctctgagacagcagaggtgaggaacttcttctatgattggt 1500  
Db 1740 GCAGGTGCTCCCTCTCTCTGAGACAGCAGAGGTGAGGACTTACTTCTATGAGTTGGT 1799  
QY 1501 cactgatgcatcagatttgcacagactgattttgacgatttttagcggaaacaaatg 1560  
Db 1800 CACGTGATGCATCAGATTGTGCACAGACTGATTTTGCACGATTTAGCGGAACAAATGTG 1859  
QY 1561 gaaactgactttagaggtccatcgcaaatgcttgaataattggttggagactgat 1620  
Db 1860 GAACATGACTTTGTAGAGTGCCATCGCAATGCTTGAAATGCTTGAAATGCTTGAGGACGTCGAT 1919  
QY 1621 tccctcgaagattgtcaaaacattataaagatggaagccctattgcagacgactgctt 1680  
Db 1920 TCCCTCCGAAGATTGTCAAAACATTATAAAGATGGAAGCCCTATTTCACAGCATCTGCTT 1979  
QY 1681 gaaacacttgttctttaggtcgtcaacacaggttcttctgacccctgcgcagattggt 1740  
Db 1980 GAAAACTTGTGCTTCTAGGCTGCTCAACACAGGCTTCTGACCCCTGCGCCAGATTGTT 2039  
QY 1741 ttgagcaaatgtgatcagctctcttaccacacacatcgctgagtgctgcaagtgaatat 1800  
Db 2040 TTGACCAAGTTGATCAGTCTCTTCAACCAACATCGCTGGATGCTGCAAGTGAATAT 2099  
QY 1801 gcaaatctgctcgaataattagagttgagctactccaggcacaataatgcagact 1860  
Db 2100 GCCAAATCTGCTCAAAATATTAGAGTTGACAGTCTTCCAGGACCAAAATATGCCACT 2159  
QY 1861 acctttgacatttggcaggggatacatgagcgaataattatgatatctttgagtgaa 1920  
Db 2160 ACCTTTGACATTTGGCAGGGGGATACCATGGCCAAATATTATGGATATCTTTGGAGTGA 2219  
QY 1921 gtattttccatgatatgttttaccagctgttttaaaaaagaggagataatgaatccagag 1980  
Db 2220 GTATTTTCCATGGATATGTTTACAGCTGTTTAAAAAAGAGGAGTAAATGAATCCAGAG 2279  
QY 1981 gttggaatgaatacagaaactaatcctgaaacctggggatctcttgacggcagtgac 2040  
Db 2280 GTTGAATGAATACAGAAACCTAATCTGAAACCTGGGGATCTCTGCACGGCATGGAC 2339  
QY 2041 atgctcacaatttctgaaacgtgagcgaacacaaacaaagcgttccctaatgtagtaggc 2100

Db 2340 ATGTCTCCACAAATTTCTGAAACGTGAGCCAAACCAAGAGCGTTCTCTATGATGAGGC 2399  
QY 2101 ctgcatgctcgtga 2115  
Db 2400 CTGCATGCTCCCGTGA 2414  
RESULT 3  
AB000170 2732 bp mRNA MAM 05-FEB-1999  
LOCUS Porcine mRNA for endopeptidase 24.16, complete cds.  
ACCESSION AB000170  
VERSION AB000170.1 GI:1783121  
KEYWORDS endopeptidase 24.16 type M3; endopeptidase 24.16 type M1.  
SOURCE Sus scrofa Adult Liver cDNA to mRNA.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE  
1 (bases 1 to 2732)  
Hirose, S.  
Direct Submission  
Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigenisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences, 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail:shirose@bio.titech.ac.jp,  
Tel:045-924-5726, Fax:045-924-5824)  
2 (sites)  
Kato, A., Sugliura, N., Saruta, Y., Hosoi, T., Yasue, H. and Hirose, S.  
Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)  
97226108  
FEATURES  
Location/Qualifiers  
1..2732  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_type="Liver"  
1..2732  
/note="corresponding to exon1.5-16 of this gene;  
endopeptidase 24.16 type 1"  
175..2289  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M: neurolysin :sBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M1"  
/protein\_id="BAAL9060.1"  
/db\_xref="GI:1783122"  
/translation="MIVRCISAAARRLHVRGSGTLLRMTLGRAMSPLQAMSSVTVDG  
RNVLRWDLSPQIKRTEELIAQTKQVDDIGMLDTEVYENCLOLADEVKYIVE  
RTMDLFPQVSSDKKEVRAASTADKRLSRSDIEKMSREDIFLRVLRKCTDGLKRP  
EARYLEKSVKMGKRLHGLPEQVQNEIKAMKRMSELCDFNKNNEDDTFLVFSKA  
ELGALPDFDLSLEKTDNNKTKITLKYFHYPMKKCCIPETRRKMEMAFNTRKEEN  
TIILQELLPLRAKAVKILGYSHADFLVEMNTAKSTHVTAFLLDLSQKLPAGEAR  
ETILNLKKECEKGEYDQKINADLHYMTQTEELKYSVDQELKYEPPIEVWTFG  
LNNIYQELLGLSFEQVTDHVNKSVTLTYTKDKATGEVLGQFVLDLYPREGYNHAA  
CFGLQGLCLPDGSRMSVAALVNFQPRAGRPSLLRHDEVTYTFEFGVHQAICA  
QTDFAFSGTNVETDFEVSQMLENNWDTLSRLSRKHYKDGSPITDLDLEKLVAS  
RLVNTGLLTLRLQIVLSKVDQSLHNTSLDAASEYAKYCTEILGVAATPGTNMPTFGH  
LAGYDQGYGYLWSEVFSMDPVSFCFKKEGINNPEVGMKYNRLILKPGGSLDGMML  
QNFLKRPNQKAFMLSRGLHAP"  
175..285  
transit\_peptide  
CDS  
244..2289  
/standard\_name="endopeptidase 24.16"  
/note="alternative product for this mRNA; oligopeptidase M  
:neurolysin :sBAP(soluble angiotensin-binding protein)  
:MEP(microsomal metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M3"  
/protein\_id="BAAL9061.1"  
/db\_xref="GI:1783123"



/translation="MTLGREAMSPLQAMSSVTVDRNVLNRWDLSPQEIQRRTTEELIAQ  
TKQVDDILMDIEVNCLOALADVEVYIVERTMLDFPOHQVSSDKVEVRAASTEA  
DKLRSRFDIEMSRREDIFLRILVKETCDLGIKIPKARRYLEKSKMGRNCLHLPEQ  
VONETKAMKKRMSLCLIDFNKLNEDDTFLVFSKALCALPDDFIDSLEKIDDDNKYKI  
TLKPHYFPVMKCIPTETRRMENAFNRCKEENTIILOELLPIURAKVALLGYSTH  
ADFLVEMTAKSTHHVTAFLDLSOKLPLGEAEREFLINLKKREKBEFYDGKIN  
AWDLHYMTQTEELKSYVDQELILKYEFPFIEVTEGLNIIYQELLGSEFQVTDHVN  
KSVLITVTKDRATGVLQFYLIDLAYPREGKNHAAFCLOPCLLPDGSRMMSVAALV  
VNFSPRAGRSLLRHSDVRVYFHEFGVHMQICAOITDFAREFSGNVETDVEVPSOM  
LENHVMWDTDSLRRLSKHYKDGSPITDDLLLEKIVASRLVNTGLLIRQIVLSKVSOLH  
TNTSLDAASEYAKVCTELGVAAATPGTNPATFGLHAGYIDGYYIYIWSSEVSDMF  
YSCFKKEGIMPEVGMKRYNLIILPGGSLDGMQLNFKLRPNOKAFIMSRGLHAP"  
BASE COUNT 826 a 545 c 641 g 720 t  
ORIGIN

Query Match 87.9%; Score 1858.2; DB 4; Length 2732;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1953; Conservative 2; Mismatches 160; Indels 0; Gaps 0;  
QY 1 atgatcgccggtgcttcttgctgtgcgaagccctccgagagtggttggttccaggatt 60  
DB 175 ATGATGTCGGTGGCTTCGGCTGGAGAGCGCTCCACAGGGTGGTGGTCCGGGATT 234  
QY 61 ttactcagaatgacgttaggaagagagtgatgtctctctccaggaatgtcttccat 120  
DB 235 TTATTGAGATGACITTAGGAGAGAGCAATGTCCTCTTCAGGCATGCTTCCTAT 294  
QY 121 actgtgctgagcaaatgttttaagatggatcttccaccagagcaaaataaaacaaga 180  
DB 295 ACTGTGATGGCAGAAATGTTTTAAGATGGATCTTTCACCAGAGCAAAATTAAGAAGA 354  
QY 181 actgaggactcattgtgcagaccacaaagggtgaagatgctgttggaatgctcggtat 240  
DB 355 ACTGAGGACCTGATTGCGCAGACCAACAAAGGTGTATGATATTGGAATGCTTGACAT 414  
QY 241 gaggagtaactacagagactgtctgcaggcactgcgcagatgagagatgaagtata 300  
DB 415 GAAGAGTAACTTATGAGAACTGTTTGAGGCACTGGCAGATGTGGAATGAGATATA 474  
QY 301 gtgaaagagaccatgtagacttcccccagcatgtatcctctgacaaagaagtacagca 360  
DB 475 GTGGAACGAACCATGCTAGACTTTCCCGAGCATGCTCCTCTGACAAAGAAGTACGAGCA 534  
QY 361 gcaagtacagaagcagacaaaagacttctctgttttgatattgagatgagcagagaga 420  
DB 535 GCAAGTACAGAGACGACAAAAGACTTTCTCGTTTTGATATTGAGATGAGCATGAGAGAA 594  
QY 421 gatatattgagagaattgttcatttacaggaaacctgtgactctggtgggaagataaacct 480  
DB 595 GATATCTTTCTGAGATTTCTCGTTTAAAGGAGACCTGTGATTTGGGGAAGATCAAGCCT 654  
QY 481 gaggccagacgatacttggaagaagtcaattaaatggggaagaagaatgggtcccatctt 540  
DB 655 GAAGCCAGAAGATACTTAGAAAAATCAGTTAAATGGGAAAAAGGAATGGGTCCCATCTT 714  
QY 541 cctggaacagtcagagaatgaaatacaatcaatgaagaaagaatgagtgagctatgtatt 600  
DB 715 CCGGAACAAGTACAGAAATCAAAAGCAATTAAGAAAGAAATGAGTACGATGATGAT 774  
QY 601 gattttaacaaaacctcaatgagatgatactctctctgtattttccaggctgaactt 660  
DB 775 GACTTCAACAAAACCTCAATGAGGACGATACCTTCTGTATTTTCCAAAGGCTGAACCT 834  
QY 661 ggtgctcttctgtagatttcttgcagtttagaaaaagacagatgatgacaaagtataaa 720  
DB 835 GGTGCTCTTCTGATGATTTTCATGACGTTTAGAAAAACAGACAGTACACAAGTATAA 894  
QY 721 attaccttaaaatataccacatttccctgctgaagaaagaatgttgatccctgaaacc 780  
DB 895 ATTACCTTAAATATCCACACTATTTTCTGTCATGAAGAAATGTTGTATCCTCGTGAACC 954  
QY 781 agaagaaggatggaatggtcttttaatacaaggtgcaagggaaaaacacccataatttg 840

DB 955 AGAAGGAAAAATGGAATGCTTTTAAATACAGGTGCAAGAGGAAAAACACCAATACTCTG 1014  
QY 841 cagcagctactccactcgcgaaccagggtggccaaactactcgtgttatagcacacatgct 900  
DB 1015 CAGGAGCTACTCCCACTCGGGCCCAAGTAGGCAAACTACTTGGGCTACAGCACACATGCT 1074  
QY 901 gacttcgtccttgaatgaactgcacagcacaagcgcgtacaagcgttcttctagat 960  
DB 1075 GACTTTGCTCTTGAGATGAACACTGCGAAGAGTACACACCATGTGACAGCCTTCTTAGAT 1134  
QY 961 gatttaagccagaagttaaaaccccttggtggaagcagaacgagagttatttttgaattg 1020  
DB 1135 GATTTAAGCCAGAAATTAACCCCTTGGGTGAGGCAAGACGAGAGTTCATTTTGAATTTG 1194  
QY 1021 agaaaaaagaatgcaaaagacagagggtttttgtaataatgaggaataatcaatgctcgtggat 1080  
DB 1195 AAAAAAAGAAATGTGAAGAGAAAGGTTTGAATATGATGGCAAAATCAATGCTGGGAT 1254  
QY 1081 ctatattactacatgactcagacagaggaactcaagtattccatagaccaagagttctctc 1140  
DB 1255 CTACATTACTATATGACTCAGACAGAAGAACCTCAAGTACTCTGTGGACCAGAGATCTCTC 1314  
QY 1141 aaggaatacttcccaattgaggtggtcactgaagggtctgtgaacacctaccaggagttg 1200  
DB 1315 AAGGAATACTTCCCAATTGAAGTGGTCACTGAAGGCTTGTCTGAACATCTATCAGGAGTTG 1374  
QY 1201 ttgggacttctatttgaacaaatgacagatgctcatgttttggacaagagtgtaacatt 1260  
DB 1375 TTGGGACTCTCTTTGAGCAAGTGACAGATGCTCATGTTTGAATAAAGTGTACCCCTG 1434  
QY 1261 tatactgtgaagataaaactacagagagaagtatttggacagttctatttggacctctat 1320  
DB 1435 TACACGGTGAAGGATAAAGCCACAGGAGAAGTATTAGSACAGTTCTACTTGGACCTCTAT 1494  
QY 1321 caaagggaagaaaaatacaatcatcgctgctcgtctcgtctcagcctgctgctcctctct 1380  
DB 1495 CCNAGGAGAGGAAAAATACAAACCATGACAGCTGCTTGGTCTCCAGGCTGGCTGCCCTCTG 1554  
QY 1381 cctgatggaagccggaatgatggcagtggtgctgctcgtgctgctgctgctcctcagcagctg 1440  
DB 1555 CCTGATGGGAGCGCATGATGCTGTGGCTGCCCTCGTGTGAACCTTCTCACAGCCACGA 1614  
QY 1441 ggaagtcgtccctctcctcagacacagagagtgagacacttacttctcagagttggt 1500  
DB 1615 GCAGGTGCTGCTCTCTCTCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGGT 1674  
QY 1501 cactgatgcatacagatttgcacagactgatttgcacgattttagcggaacaaatggt 1560  
DB 1675 CAGGTCAATGCATCAGATTTGTGGCAGACTGACTTTTGCAGGATTTAGTGAACAAATGTG 1734  
QY 1561 gaaactgactttagagtgccatgcgaatgctgcaaatgctgcaaatggtggtggtggtggtggt 1620  
DB 1735 GAAACTGACTTTGTAGAGGTGCCATCACAATGCTTGAACCTGGGTGTGGGACACTGAT 1794  
QY 1621 tccctccgaagattgtcaaaacattataaagatggaagccctatttcagacagctcgtt 1680  
DB 1795 TCCTCCGAAGGCTGTCAAAACATTTAAAGATGGAAGCCCTATTACAGATGATCTGCTT 1854  
QY 1681 gaaaaacttgttctcttaggtggttcaacacaggtctctctgacctggtggtggtggtggtggt 1740  
DB 1855 GAAAACTTGTGCTCTAGACTGGTCAACACAGGTCTCTCTGACCTTACGCCAGATGTTT 1914  
QY 1741 ttgagcaaatgtgactgtctcttctacacacacatcgctggtgactgctgcaatgaaat 1800  
DB 1915 TTGAGCAAAAGTTGATCAGTCTCTCCACACCAACACTTCGTGCGTGGATGCTCAAGTGAATAT 1974  
QY 1801 gccaaatactgctcagaaataattaggagttgcaagctactctccagggcaaaataatgcagct 1860  
DB 1975 GCCAAATCTGCACAGAAATTTTAGTGTGTGACGTACTCCAGGCACAAATATGCCAGCT 2034  
QY 1861 accttggacattgagcagggggagatcgatggccaataattatgatatcttggagtgaa 1920



Db 998 CAGGAGCTACTCCACTCGGGGCCAAAGTAGCCAAACTACTGGGCTACAGCACACATGCT 1057  
Qy 901 gacttgccttgaaatgaactgcaaaagacacaaagccgcgtaaacacagcytttctagat 960  
Db 1058 GACTTTGTCTTGTAGATGAACACTGCGAAGAGTACACACATGATGACAGCCCTTCTAGAT 1117  
Qy 961 gatttaagccagaagttaaacccttgggtgaagcagaacgagagtttatttgaatttg 1020  
Db 1118 GATTTAAGCCAGAAATTAACCCCTTGGGTGAGCGCAGACGAGAGTTTCATTTGAATTTG 1177  
Qy 1021 agaaaaaggaatgcaaaagcaggggttttgaataatgatgggaaaaatcaatgcctgggat 1080  
Db 1178 AAAAAAAGAAATGTCAGAGAAAGCTTTTGAATATGATGGCAAAATCAATGCTGGGAT 1237  
Qy 1081 ctatattactacatgactcagacagaggaaactcaaatgatttccatagaccagaagattcctc 1140  
Db 1238 CTACATATTATATGACTCAGACAGAAAGAACTCAAGTACTCTGTGGACCAAGAGATCCCTC 1297  
Qy 1141 aaggaatacttcccaattgaggtggctcactgaagcgttctgtgaacacacctaccaggagttg 1200  
Db 1298 AAGGAATATTCCTCAATTTGAAGTGGTCACTGAAGGCTTGTGACACATCTATCAGGAGTTG 1357  
Qy 1201 tggggacttcttattgaacaaatgacagatgctcatgttttggaaacagagtggtacactt 1260  
Db 1358 TTGGGACTCTCCTTTGAGCAAGTGACAGATGCTCATGTTTGGAAATAAAAGTGTTACCCCTG 1417  
Qy 1261 tatactgtaagagataaagctacagagagaagtattgggacaggttctatttggacctctat 1320  
Db 1418 TACACGGTGAAGGATAAAGCCACAGAGAAAGTATTAGGACAGTTCCTACTTGGACCTCTAT 1477  
Qy 1321 ccaaggaagaaaaatcacatcatcgctgcttgcgtctcagctctccagcgtgctgctctctt 1380  
Db 1478 CCAAGGAGAGAAAAATACAAACATGACGCTGCTTGGTCTCCAGGCTGGCTGCCCTCCG 1537  
Qy 1381 cctgatggaagccggatgatggcagtggtgctcctcggtggaacttctcacagccagtg 1440  
Db 1538 CCGTATGGGAGCGCATGATGCTGTGGTGCCTCGTGGTGAACCTTCACAGGCCACGA 1597  
Qy 1441 gcaagtcgctcctctctctgagacacgacgagagtgtaggacttaacttcatgagttggt 1500  
Db 1598 GCAGGTGCGCCCTCTCTCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGTT 1657  
Qy 1501 cagctgagtcacagattgtgacagactgatttgcagcagatttagcggaacaaatgtg 1560  
Db 1658 CAGTCATGCATCAGATTGTGCGCAGACTGACTTTGCAGGATTTAGTGAACAAATGTG 1717  
Qy 1561 gaaactgacttttagagtgccatgcgaatgcttgaataattgggtgtgggacgtcgat 1620  
Db 1718 GAAACTGACTTTGTAGAGTGCCATCACAAATGCTTTGAAAACTGGGTGTGGGACACTGAT 1777  
Qy 1621 tccctccgaagattgtcaaaacattataagatggaagccctatttcagacgactgctt 1680  
Db 1778 TCCTCCGAAGGCTGTCAAAACATTATAAGATGGAAGGCCCTATTACAGATGATCTGCTT 1837  
Qy 1681 gaaaaacttgcttcttagctgtggtcaacacaggtctcttcgacctgcccagattggt 1740  
Db 1838 GAAAAACTTGCTCTAGACTGGTCAACACAGGCTCTCTGACCTTACGCCAGATTGTT 1897  
Qy 1741 ttgagcaagttgatcagttcttctacacacacacatcgctgtagtgcgaatgaaat 1800  
Db 1898 TTGAGCAAAAGTTGATCAGTCTCTCCACACCAACACTTCGCTGGATGCTGCAAGTGAATAT 1957  
Qy 1801 gcaaaactgctcagaataattagagattgagctactccaggaacaataatgcagct 1860  
Db 1958 GCCAAATCTGCACAGAAATTTTAGTGTGTGAGCTTCTCCAGGCAAAATATGCCAGCT 2017  
Qy 1861 acctttggacatttggcaggggatacagtgcccaataattatgatattctttggagtgaa 1920  
Db 2018 ACTTTTGGACACTTGGCAGGGGATATGATGGCCATATATTATGATATCTTTGGAGTGA 2077  
Qy 1921 gtattttccatgatattgtttacagctgttttataaaaaagaaggaataatgaatccagag 1980  
Db 2078 GTCTTTTCCATGGACATGTTTATTAGTCTGTTTAAAAAAGAGGAGGATTAATGAATCCAGAG 2137

Qy 1981 gttggaataacatacaaaacctaatactgaaacctggggatctctgacgcacatgagc 2040  
Db 2138 GTTGGAAATGAATACAGAAACCTTAATCTCGAAACCTGGGGGATCGCTGGACGATGGAC 2197  
Qy 2041 atgctccacaatttcttgaacacgtgagcgaacacacaaacgcttccctaatgagtagagc 2100  
Db 2198 ATGCTCCAAATTTCTTGAACACGTGAGCCAAACCAACCAACGCTTCTCTAATGAGTAGAGGC 2257  
Qy 2101 ctgcatgctctcgtga 2115  
Db 2258 CTGCATGCTCCGTAA 2272

RESULT 5  
AB000172 2810 bp mRNA MAM - 05-FEB-1999  
LOCUS Porcine mRNA for endopeptidase 24.16, complete cds.  
DEFINITION  
ACCESSION AB000172  
VERSION AB000172.1 GI:1783126  
KEYWORDS endopeptidase 24.16 type M2; endopeptidase 24.16 type M3.  
SOURCE Sus scrofa Adult Liver cDNA to mRNA.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE  
AUTHORS Hirose,S.  
TITLE Direct Submission  
JOURNAL  
SUBMITTED (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigehisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail:shirose@bio.titech.ac.jp,  
Tel:045-924-5726, Fax:045-924-5824)  
2 (sites)  
REFERENCE  
AUTHORS Kato,A., Saruta,N., Hosoi,T., Yasue,H. and Hirose,S.  
TITLE Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
JOURNAL J. Biol. Chem. 272 (24), 15313-15322 (1997)  
MEDLINE 97326108  
FEATURES  
source  
1. .2810  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_type="Liver"  
1. .2810  
/note="corresponding to exon2,5-16 of this gene;  
endopeptidase 24.16 type 2"  
130. .2367  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M :neurolysin :SBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M2"  
/protein\_id="BAA19063.1"  
/db\_xref="GI:1783127"  
/translation="WVPEGLHRELGLATFSSAPLGGHPFFVWDCLSCKOGDWSQA  
RPTNAERSSGGSGILLRMTLGRAMSPLOAMSSYTVDRNLKRWLSDPEQIKRRRT  
ELIAQTQVDDIGMLDIEYENCLOALADVEYKIVERTMLDLPQHVSDKKEVR  
AASADKRLSFDIEMSNREDIFLRVRLKETDGLKIKPEARRYLEKSRVKKRNG  
LHLPOVONEIKAMKRMSELCTDFKNLNEEDTFLVFSKALGALPDDFIDISLEKTD  
DNKYKTLKYPHVFPVKKCCIPETRRKMEMAFNCKEENTILLOELLPLRAKVAL  
LGYSHADVFLEMTAKSTHHVTAFLDLSQKLPLGEAREFELNKKKEBEKFE  
YDGNINADLHYMTQTEELKYSDQEIILKEYFPPIEVYEGLLNIQELLGUSFEQVT  
DAHVNKSVTLYTKVDKCEVLGQFYLDLYPREGYNHAACFTGLOPGCLLPDSRRM  
SVAALVNPSPRAGPSLLRSHDEVITYPEFHVHQICAOQDFARFSGTNNVETDV  
EYPSOMLENVNDTSLRLSKHYKDGSPITDLDLEKLVASRLVNTGLTLQIVLSK  
VQOSLTNTSLDAASEYAKYCTEILGVAATPGTNMPATGCHLAGVDGQYCYLASEV  
FSDMFYSCFKKEGIMNPVGMKIRNLILKPGSLDGMQLQNLFLAKREPQKAFLMR  
GLHAP"  
322. .2367  
/standard\_name="endopeptidase 24.16"

```
/note="alternative product for this mRNA; oligopeptidase M
:neurolysin :sBAP(soluble angiotensin-binding protein)
:MEP(mitochondrial metalloendopeptidase)"
/codon_start=1
/product="endopeptidase 24.16 type M3"
/protein_id="BAA19064.1"
/db_xref="GI:1783128"
/translation="MTLGREAMSPLOAMSSYTVDRNVLKRWLSPQLKRRTEELIAQ
TKQVDDIDMLDIEEDVENCLOALADVEVKYIVERTMLDFPQHVSSDKVRAASTEA
DKRLSRFDIEMSRDIEDIFLRVLKCTDLGKIKPEARRYLEKSVKMGKRNGLHLPEQ
VQNETKAMKRMSELCDIPNKNLNEDDTFLVSKAELGALPDDPIDSLEKTDNNKYKI
TLKLYFPVWKCKCIPETRRKMEAFNCRKEENTIIQLLELLPURAVALKLGYSTH
ADFLVEMTAKTTHVITADLDLSOKLPLGEAREFLLNKKKECEKEKFEYDQKIN
AWDLHYMTQTEELKYSVDQELKYEFPFIEVTEGLNIIYQELLGLSPQEVQTDAAVW
KSVLTLYTKDKATGEVLFQFDLYLREGKYNHAAFCGLPCLLPDGRMMSVAALV
VNFSPRAGRSLRLRHDKSVRTYEFHEFGVHMQICAOITDFARESGTNETDFEVPSON
LENWWDPSLRLLKHKYKDGSPITDDLLLEKLIVASRLVNTGLLRLQVLISKVQSLH
TNTSLDAASEYAKYKTELLGVAATPGTNMPATFGLHAGGYDQGYIYIWSVFSMDMF
YSCFKKEGIMNPEYGMKYRNLILKPGGSLDGMQLNLFKREPNOKAFMLMSRLHAP"
BASE COUNT      852 a  558 c  652 g  748 t
ORIGIN

Query Match      86.5%; Score 1829.8; DB 4; Length 2810;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1933; Conservative 2; Mismatches 174; Indels 0; Gaps 0;

QY  7  gccggtgccttttgcgtgcgaagcctccgaagttggtggtccagagatttactc 66
DB 259  GCCGCCCCAAAACGAATCCAGAAAGACGGTCGGGGTGTGGTTCGGGGATTTATG 318
QY  67  agaagcgttaggaagaagtgatgtctctcttcaggcaatgtcttccatactgtg 126
DB 319  ACAATGACCTTTAGGGAGAGAACCAATGCTCTCTTCAGGCAATGCTTCCTATACTGTG 378
QY  127  gctgcagaaatgcttttaagatggaattcttccagagcaaatataaacaagaactgag 186
DB 379  GATGGCAGAAATGTTTAAAGATGGGATCTTTCACCAGAGCAAAATTAAGAAGAACTGAG 438
QY  187  gactcattgtcacacaaacaggttacgatgtctgttggaatcctcgttattgagaa 246
DB 439  GAGCTGATGGCCGACACCAACAGGTGATGATGATATGGAAATGCTTGACATTAAGAAG 498
QY  247  gtaacttacgagaactgtctgcaggcaactggcagatgtagaagtaagatatagtgaa 306
DB 499  GTAACCTATGAGAACTGTTTGACGCACTGGCAGATGTGGAACTGAAGTATATAGTGAA 558
QY  307  aggaacatgctagactttcccagcatgtatcctctgcacaaagaagtcagagcagaagt 366
DB 559  CGAACCATGCTAGACTTTCCCCAGCATGTCTCCTCTGACAAAGAAGTACGAGCAGCAAGT 618
QY  367  acagaacagacaaaagacttctcgttttgatattgagatgagatgagagagatatata 426
DB 619  ACAGAAGCAGACAAAAGACTTTCCTGTTTGATATGAGATGAGATGAGAGAAGATATC 678
QY  427  tttagaagaattgttcattacagaaacacctgtgatctggggaagataaaacctgagcc 486
DB 679  TTCTGAGAAATGTTTCGTTTAAAGAGAGACCTGTGATTTGGGGAAAGATCAAGCCTGAAGCC 738
QY  487  agacgactacttggaagaagtcaattaaaaatggggaagaagaatgggctccatctctgaa 546
DB 739  AGAAGATACTTAGAAAAATACAGTTAAAATGGGAAAAAGAAATGGGCTCCATCTCCGGAA 798
QY  547  caagtacagaatgaatacaatcaatgaagaaagaatgagtgaactgattgatttt 606
DB 799  CAAGTACAGAAATGAATCAAGCAATGAAGAAAGAAATGAGTGAGCTATGATTAAGCTTC 858
QY  607  aacaaaaacctcaatgagatgatacccttctgttatctccaaagctggaacttggtgct 666
DB 859  AACAAAAACCTCAATGAGGACGATACCTTCCTGTGATTTTCCAAAGCTGAACCTGGTGTCT 918
QY  667  cttcctgatgatttcattgacagtttagaagaacagatgatgacaagtataaaattacc 726
```

```
DB 919  CTTCTGTGATGATTTCATTGACAGTTTAGAAAAAGACAGATGACAAAGTATATAATTACC 978
QY  727  ttaaatatccacacatttccctgtcatgaagaatgttgatccctgaaaccagaaga 786
DB 979  TTAANAATCCACATATTTTCTGTGATGAAGAAATGTTGTATCCCTGAACACCAAGG 1038
QY  787  aggatggaatggcttttaatacaagggtgcaagaggaagaaaaacaccataatttgcagag 846
DB 1039  AAAATGGAAATGGCTTTTAATACGAGGTCAAAGAGGAAACACCATATATCTCGCAGGAG 1098
QY  847  ctactcccaactgcgaaccaaggtggccaaactactcggttatagcacacatcgtgacttc 906
DB 1099  CTACTCCCACTCGGGCCCAAGTACCCAAACTACTGGGTACGACACACATCTGCTGATTT 1158
QY  907  gctcctgaaatgaacactgcgaagagcacaagccgtaacagcvtcttagatattta 966
DB 1159  GTCCCTTGAGATGAACACTGCGGAAGAGTACACACCATGTGACAGCCCTTCTAGATGATTTA 1218
QY  967  agccagaagttaaaaccccttgggtgaaagcagacagagatttatttgaatttgaagaaa 1026
DB 1219  AGCCAGAAATTAACCCCTTGGGTGAGCGAAGACGAGAGTTCATTTGAATTTGAAGAAA 1278
QY  1027  aaggaatgcaagacagaggggttttgaatatatgatgggaaaaatacaatgcctgggatatata 1086
DB 1279  AAGGAATGTGAAGAGAAAGGTTTGAATATGATGCAAAATCAATCCCTGGGATCTACAT 1338
QY  1087  tactacatgactcagacagaggaactcaagtattcctatagaccagagttcctcaagaaa 1146
DB 1339  TACTATATGACTCAGACAGAGAACTCAAGTACTCTGTGGACCAAGAGATCTCTCAAGGAA 1398
QY  1147  tacttcccaattgaggtggtcactgaaaggcttgcgtgaacacactaccagaggttgttgga 1206
DB 1399  TACTTCCCAATTTGAAGTGGTCACTGAAGGCTTGGCTGAACATCTATCAGGAGTGTGTGGGA 1458
QY  1207  cttcatttgaacaaatgacagatgctcatgtttggaacaaagttgtacactttatact 1266
DB 1459  CTCCTCTTTGAGCAAGTACAGATGCTCATGTTTGGAAATGTTTACCCCTGTACAGG 1518
QY  1267  gtgaagataaagctcacaggaagatttggacagttctatttggacctctatccaagg 1326
DB 1519  GTGAAGGATAAAGCCACAGGAGAAATATTAGACAGATTCTTACTTGGACCTCTATCCAAGG 1578
QY  1327  gaagaaaaatacaatcgtcggtcgtctccagcgtggtcgtcgttctgctgctgat 1386
DB 1579  GAAGGAANAATACAACCATGACGCTGCTTGGTCTCCAGCCTGGCTGCTCTGCTGCTGAT 1638
QY  1387  ggaagccggaatgtaggcagtggtgctcctcgttggtaacttctcagaccagtgaggt 1446
DB 1639  GGGAGCCGATGATGCTGTGTGCTGCCCTCGCTGGTGAACCTTCTCACAGCCACGAGCAGGT 1698
QY  1447  cgtccctctcctgagacagacagaggtgaggttacttctcatgatttgggtcacgtg 1506
DB 1699  CGTCCCTCTCTCCTGAGACACGATGAGGTGAGGACTTACTTCCATGAATTTGGTCAAGTTC 1758
QY  1507  atgcatcagatttgcacagactgatttgcacgatttagcggaacaaaatgtggaact 1566
DB 1759  ATGATCAGATTTGTGGCCAGACTGACTTTGCACGATTTAGTGGAAACAAATGTGGAACAT 1818
QY  1567  gacttggtagagggtccatcgcaaatgttgaaaattgggtgtagggcgtcgattccctc 1626
DB 1819  GACTTGTGAGAGGTGCCATCAAAATGCTTGAATACTGGGTGGGACACTGATTTCCCTC 1878
QY  1627  cgaagatgtcaaacattataaagatggaagccctatttgcagacgactctgcttgaataa 1686
DB 1879  CGAAGCTGTCAAACATATTAAGATGGAAGCCCTATTTACAGATGATCTGCTTGAAGAAA 1938
QY  1687  cttgtgcttctaggtgttcaacacaggttcttgcacctgcccagattgttttgagc 1746
DB 1939  CTTGTGCTTCTAGACTGGTCAACACAGGCTCTCCTGACCTTACGCCAGATGTTTGTGAGC 1998
QY  1747  aaagttgatgctcttctcataccacaacatcgctggatgctggaagtgaataatgccaaa 1806
DB 1999  AAAGTTGATGCTCTCTCCACACCAACCACTTCGCTGGATGTGCAAGTGAATATGCCAAA 2058
```



Db 1025 AAACACTGCGGTACAGCACACATGCTGACCTTGTCTTGTGAGTGAACACTGCGAAGAGT 1084  
Qy 934 acaagccgcgttaacagcycttctagatgatatttaagcagaagagtttaaaacccctgggtgaa 993  
Db 1085 ACACACCATGTCAGACCCCTTCTAGATGATTTAAGCCAGAAATTTAAACCCCTGGGTGAG 1144  
Qy 994 gcagacagagttatttttaatttgaagaaaggaatcaagacagaggggttttga 1053  
Db 1145 GCAGAACGAGAGTTCATTTTGAATTTGAAGAAAGAGGATGTGAAGAAAGGTTTGA 1204  
Qy 1034 tatgtgggaaaaatcaatgcctgggatctattactactacgactcagacagagaaactc 1113  
Db 1205 TATGATGCCAAAATCAATGCGGTGATCTACATATATGACTCAGACAGAAAGAACTC 1264  
Qy 1114 aagattccatagacccaagagttcctcaaggaatacttcccaattgaagtggtcaactgaa 1173  
Db 1265 AAGTACTCTGTGGACCAAGAGATCCTCAAGAAATACTTTCCCAATTTGAAGTGGTCACTGAA 1324  
Qy 1174 ggtctgtgaacacctaccagagtggttggaactttcatttgaacaaatgacagatgct 1233  
Db 1325 GCCTGCTGACATCTATCAGAGTGTGTGGACTCTCCTTTGACCAAGTCAAGATGCT 1384  
Qy 1234 catgtttggaacaagagtggttacactttatactgtgaaggataaagctcacaggagaagta 1293  
Db 1385 CATGTTTGAATAAAAGTGTACCTGTACACGGTGAAGGATAAAGCCACAGGAGAAGTA 1444  
Qy 1294 ttgggacagttctatttgacactctatccaaagggaagaaatacaatcatctgcgcctgc 1353  
Db 1445 TTAGGACAGTTTCTACTTGGACCTCTATCCAAAGGAAGGAAATACAACTATGACAGCTGC 1504  
Qy 1354 ttcggtcccaagcctgcctctctgctgctgacgtgaagccgagatgagcagtgctgc 1413  
Db 1505 TTTGGTCTCCAGCCTGGCTGCTGCTGCTGATGGGAGCCGACGATGCTGTGGCTGCC 1564  
Qy 1414 ctgctgtggaactctcagccagtgccagtgctgctcctctcctgagacacagcagag 1473  
Db 1565 CTCGTGTGTGAATCTTCACAGCCACGACGAGTGTCTCTCTCTCTGAGACACGATGAG 1624  
Qy 1474 gtgaggaacttcttcacatgatttggtcaagtgatgacatgatttgacagactgat 1533  
Db 1625 GTGAGGACTTACTTCCATGAATTTGGTTCACCTCATGTCATGATGTTGTGCGACAGCTGAC 1684  
Qy 1534 ttgcacgatttagcggaacaaatgtgaaactgactttgtagagtgccatcgcaaatg 1593  
Db 1685 TTTGCACGATTTAGTGGAAACAAATGTGGAACACTGACTTTGTAGAGGTGCCATCACAATG 1744  
Qy 1594 cttgaaaattgggtgtggagcgtgattccctccgaagattgtcaaaacattataaagat 1653  
Db 1745 CTTGAAAACCTGGGTGTGGACACTGATTCCTCCGAAAGGCTGTCAAAACATTTATAAAGAT 1804  
Qy 1654 ggaagccctattgcagacgactctgtgaaacttggctcttagctgtgctgacacaca 1713  
Db 1805 GGAAGCCCTATTACAGATGATCTGCTTGAAAACTTGTGTCTCTAGACTGGTCAACACA 1864  
Qy 1714 ggtctctgacccctgcgcagattgttttgagcaaatgtatcagctctctcattaccaac 1773  
Db 1865 GGTCTCTGACCTTACGCCAGATTTGTTGAGCAAGTTGATGATGATCTCTCCACCAAC 1924  
Qy 1774 acatcgctggatgctgcaagtgaatatgccaaaactactgctcagaaaattattaggagtgc 1833  
Db 1925 ACTTCGCTGGATGCTGCAAGTGAATATGCCAAATACTGCACAGAAATTTTAGGTGTGCA 1984  
Qy 1834 gctactcagcagcaaatatgcagctacccttggacatttggacatttggcaggggatacagatggc 1893  
Db 1985 GCTACTCCAGGCAAAATATGCCAGCTACTTTTGGACACTTTGGCAGGGGATATGATGGC 2044  
Qy 1894 caattatgatatcttggagtgaaagtatttccatggatatttccatgagctgtttt 1953  
Db 2045 CAATATTATGATATCTTGGAGTGAAGTCTTTTCCATGGACATGTTTATAGCTGTTT 2104  
Qy 1954 aaaaaaaggggataatgaatccagaggttggatgaatacagaaacctaactcctgaa 2013

Db 2105 AAAAAAGAGGGATAATGAATCCAGAGTTTGAATGAATAACAGAAACCTAATCTGAAA 2164  
Qy 2014 cctggggatctctggacgcgcatgacatgctccacatttcttgaacgtgagcgaac 2073  
Db 2165 CCTGGGATCGCTGGAGCGCATGACATGCTCCAAATTTCTTGAACAGTGAGCCAAAC 2224  
Qy 2074 caaaaagcgttccaaatgagtagagggcctgcatgctcgtcgtga 2115  
Db 2225 CAAAAGGTTCTTAATGAGTAGAGGCTGCGTAA 2266

## RESULT 7

AB000173 LOCUS AB000173 2862 bp mRNA MAM 05-FEB-1999  
DEFINITION Porcine mRNA for endopeptidase 24.16, complete cds.  
ACCESSION AB000173  
VERSION AB000173.1 GI:1783129  
KEYWORDS endopeptidase 24.16 type M3.  
SOURCE Sus scrofa Adult tissue\_lib:Liver cdna to mRNA.

## ORGANISM

Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
Hirose, S.

## REFERENCE

1 (bases 1 to 2862)  
Direct Submission  
Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.  
Shigehisa Hirose, Tokyo Institute of Technology, Department of  
Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama,  
Kanagawa 226-8501, Japan (E-mail: shirose@bio.titech.ac.jp,  
Tel:045-924-5726, Fax:045-924-5824)

## REFERENCE

2 (sites)  
Kato, A., Sugiyama, N., Saruta, Y., Hosoiri, T., Yasue, H. and Hirose, S.  
Targeting of endopeptidase 24.16 to different subcellular  
compartments by alternative promoter usage  
J. Biol. Chem. 272 (24), 15313-15322 (1997)

## FEATURES

Location/Qualifiers  
1..2862  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/dev\_stage="Adult"  
/tissue\_lib="Liver"  
1..2862  
/note="corresponding to exon2.4.5-16 of this gene;  
endopeptidase 24.16 type 2"  
374..2419  
/standard\_name="endopeptidase 24.16"  
/note="oligopeptidase M :neurolysin :sBAP(soluble  
angiotensin-binding protein) :MEP(microsomal  
metalloendopeptidase)"  
/codon\_start=1  
/product="endopeptidase 24.16 type M3"  
/protein\_id="BAI19065.1"  
/db\_xref="GI:1783130"  
/translation="MTLGREAMSPLQAMSSYTVGDRNVLRLWDLSPQIKRRTBELIAQ  
TKGVYDIDGLMDEEVTYENCLADVEVKYIVERTMLDFPQHVSSDKVRAASTEA  
DKLSRFDIEMSRMEDIFLVRLKCTDLGKIKPEARVLEKSRKMGKNGHLHPQ  
VONEIKAMKRMSELCTIDFNKLNEDDTFLVSKAELGALPDDFIDSLKTDNDKVKYI  
TLKYPHYPMKKCCIPETRRKMEMAFNTRCKEENTIIQLLEPLRAKVAKLLGYSTH  
ADVLMENTAKSTHHVTLDLUSQKLPLGEAREFILLNKKKECEKEGFEDGKIN  
AMDHYMTQTTELKYSDQELIKYFPIEVVTGELINIQELLGLSFEQVTDHVVN  
KSVTLTVKDKATGEVLGYLDLYPREGYNHACFGLOPGLLPDGRMSVAALY  
VNFSPRAGRPSLLRHDEVYTFEFGVHMHQCAQTFARFGCTNNVETDFEVDPSOM  
LENWVNDTSLRSLSKHYKDGSPITDLDLEKLVASRLVNTGLTLRLQIVLSKVDOSLH  
TNTSLDAASEYAKYCTEILGVAATPGTNMPATFGLHLAGGYDGYQYGLWSEVSMDF  
YSCFKKEGINPNPVGKMYRNLIILKPGSGSDGMDMLQNLFLKREPDKAFLMSRGLHAP"  
BASE COUNT 866 a 568 c 656 g 772 t

ORIGIN  
Query Match 86.4%; Score 1828.4; DB 4; Length 2862;  
Best Local Similarity 92.3%; Pred. No. 0;  
Matches 1922; Conservative 2; Mismatches 158; Indels 0; Gaps 0;



QY 34 ctccgcagagtggtggttccaggattttactcacaatgacatgacgttaggaagagaagtatg 93  
 DB 338 CACAATCTGGTTGGTGTCCGGGATTTTATGAGAAAGACTTTAGGGAGAGAAACAATG 397  
 QY 94 tctctcttcaggcaatgctcttcctactactggtggtgagaaaatgttttaagatgggat 153  
 DB 398 TCTCCTCTTCAGGCAATGCTTCTCTATCTGTGGATGGCAGAAATGTTTTAAGATGGGAT 457  
 QY 154 ctttcaccagagcaaatataaacagaactgaggagctcatgtgcagacccaaacaggtg 213  
 DB 458 CTTTCAACCAGAGCAAAATTAAGAGAGAACTGAGGAGCTGATTGCGCAGACCAACAAGGTG 517  
 QY 214 tacgatgctgttggaatgctgcgtttaggaagtaacttacgagaaactgtctgcaggca 273  
 DB 518 TATGATGATATTGGGAATGCTTACATTTGAGAGGTAACTTATGAGAACTGTTTGAGGCA 577  
 QY 274 ctggcagatgagaagtaagatatagtggaagagaccatgctagactttcccagcat 333  
 DB 578 CTGGCAGATGTGGAAGTGAAGTATATAGTGAACGAACCATGCTAGACTTTCCCCAGCAT 637  
 QY 334 gtatcctctgacaagaagtaacgacgacaagtacagaagcagacaaagaactttctcat 393  
 DB 638 GTCTCCTCTGACAAAGAGTACGACGACGCAAGTACAGAAAGCAGACAAAGAGACTTCTCT 697  
 QY 394 ttgtgattgagatgagcatgagagagatatatttgagagaattgttcatctacaggaa 453  
 DB 698 TTTGATATTGAGATGACATGACATGAGAGAGATATCTTCTGAGAAATGTTCTTAAAGGAG 757  
 QY 454 acctgtgatctgggaagataaaaacctgagccagacgatacttggaaaagtcatttaa 513  
 DB 758 ACCTGTGATTTGGGAAGATCAAGCTCGAAGCCAGAAGATCTTAGAAAAATCAGTTAAA 817  
 QY 514 atggggaaaagaattgggtccatctctctgaaacaggtacagaatgaaatcaaatcaatg 573  
 DB 818 ATGGGAAAAGAAATGGGCTCCATCTTCGGAAACAAGTACAGAATGAATCAAAACAATG 877  
 QY 574 aagaaaagaatgagtgagctatgtattgattttaacaaaaacctcaatgagatgatacc 633  
 DB 878 AGAAAAAGAAATGAGTGAGTATGTATGTACTTACAAAAAACTCAATGAGGACGATACC 937  
 QY 634 ttcttgattttccaaggctgaacttggtgctctctctgattgatttcattgacagtta 693  
 DB 938 TTCCCTGTATTTTCCAAGGCTGAACCTTGCTGCTCTTCTCTGATGATTTTCATTGACATT 997  
 QY 694 gaaagacagatgatacagaatataaaatctacttaaaataccacactatttcctgtc 753  
 DB 998 GAAAAGACAGATGACAACAAGTATAAAATTTACCTTAAAAATATCCACACTATTTTCTGTC 1057  
 QY 754 atgaagaatgtgtatccctgaaaccagagaagagatggaatggcttttaatacaagg 813  
 DB 1058 ATGAAGAAATGTTGTATCCCTGAAACCCAGAGAAAGTGAATGGCTTTTAAATACGAGG 1117  
 QY 814 tgcaagaggaataacacataaatttgacagcgtactcccactgcgaaccaagtgcc 873  
 DB 1118 TGCAAGAGAGAAAACACCATATATCTGAGGAGCTACTCCCCTGCGGCGCCAAAGTAGCC 1177  
 QY 874 aaactactggttatgacacacatgctgactctgctctcttgaatgaacactgcaaaagac 933  
 DB 1178 AAATCTAGGCTACAGCACACATGCTGACTTTGCTTTGAGATGAACATCGGAAGAT 1237  
 QY 934 acaagccgataacagcyttcttagatgatttaagccagaagtataaacctctgggtgaa 993  
 DB 1238 ACACACCATGTGACAGCCCTTCTAGATGATTTTAAAGCAGAAATTTAAACCTTTGGGTGAG 1297  
 QY 994 gcagacgaggttattttgaattgaagaaaaggaatgcaagacaggggttttga 1053  
 DB 1298 GCAGAACGAGAGTTCATTTTGAATTTGAAGAAAAGGAATGTGAAGAGAAAGTTTGA 1357  
 QY 1054 tatgatgggaaaatacaatgcctgggtatctatatctactactgactcagacagaggaactc 1113  
 DB 1358 TATGATGGCAAAATCAATGCCCTGGGATCTACATTACTATATGACTCAGACAGAAACATC 1417

QY 1114 aagtattccatagaccaaagagttctcaaggaaatacttcccattgaggtggtcaactgaa 1173  
 DB 1418 AAGTACTCTGTGGACCAAGAGATCTCAAGGAATACTTCCCAATTTGAAGTGTCACTGAA 1477  
 QY 1174 ggtctgtgaacacatcacacagaggttgggtggaacttcaatttgaacaaatgacagatgc 1233  
 DB 1478 GGCTTGTCTGAACATCTATCAGGAGTTGTGGGACTCTCTCTTTGAGCAAGTGACAGATGCT 1537  
 QY 1234 catgtttggaaacaagagtggttacactttatactgtgaagataaaagctacagagagaaga 1293  
 DB 1538 CATGTTTGGAAATTAAGTGTTTACCTGTATACCGGTGAAGGATAAAGCCACAGAGAAGTA 1597  
 QY 1294 ttggacaggttctatttgacacctatccaagggaaggaatacaaatcatcgccctgc 1353  
 DB 1598 TTAGACAGTCTTACTTGGACCTCTATCCAGGAGGAGAAATACACCATCGAGCTGC 1657  
 QY 1354 ttcggtctccagcctgggtgctcttgcctgatggaagccggatgatggcagtggtgcgc 1413  
 DB 1658 TTTGGTCTCCAGCCTGGCTGCTCTGCTGATGGAGCCGATGATGCTGTGGGTGCC 1717  
 QY 1414 ctgctgggtgaacttctcacagccagtggaagctgtccctctctctcagacagacagag 1473  
 DB 1718 CTCGTGGTGAACCTTCTCAGCCACGAGCAGGTGCTCTCTCTCTGAGACCGATGAG 1777  
 QY 1474 gtgagacttactttcatgagtttggtcaacgtgacatcagatgttgacacagactgat 1533  
 DB 1778 CTGAGGACTTACTTCCATGAATTTGGTCACTGTCATGCATCAGATTGTGCGGACAGCTGAC 1837  
 QY 1534 ttgcaacgatttagcgggaacaaatgtggaactgactttgttagagtgccatgcgcaatg 1593  
 DB 1838 TTTGACGACTTTAGTGGAAACAAATGTGGAACACTGACTTTGTAGAGGTGCCATCACAAATG 1897  
 QY 1594 cttgaaaattgggtgggacgtgattccctccgaagattgtcaaaacattataaagat 1653  
 DB 1898 CTTGAAAACCTGGGTGGGACACTGATTCCTCCGAAGCTGTCAAAACATTATAAAGAT 1957  
 QY 1654 ggaagccctattgcagaacgatctgtgaaaactgttgcctctaggtgcggtcaaacaca 1713  
 DB 1958 GGAAGCCCTATTACAGATGATCTGCTTGAAGAACTTGTGCTTCTAGACTGCTCAACACA 2017  
 QY 1714 ggtctctgacccctgcgcagattgtttgagcaaatgttgatcagtcctctctcacaac 1773  
 DB 2018 GGTCTCTGACCTTACGCCAGATTTGTTTGAAGTGTGATCAGTCTCTCCACACCAAC 2077  
 QY 1774 acatgcgtgagtgctgcgaagtgaataatgcaaaactgctcagaataattagagattgca 1833  
 DB 2078 ACTTCGTGATGCTGCAAGTGAATATGCAAAATCTGCACAGAAATTTTAGGTGTGCA 2137  
 QY 1834 gctactccagcacaaaatatgccactcttgacatttgacagggggagatacagatgc 1893  
 DB 2138 GCTACTCCAGGCACAAATATGCCACTTCTTTGGACACTTGGCAGGGGATATGATGCG 2197  
 QY 1894 caatatattgatatctttggagtggaagtattttccatggtatgttttacagctgtttt 1953  
 DB 2198 CAATATTATGGATATCTTTGGAGTGAAGTCTTTTCCATCGACATGTTTATAGTGTCTTT 2257  
 QY 1954 aaaaagaaggggataatgaatccagaggttggaatgaaatacagaaaacctaactccatgaa 2013  
 DB 2258 AAAAAAGAGGGATAATGAATCCAGAGGTGGAATGAAATACAGAAACCTTAATCTCTGAAA 2317  
 QY 2014 cctgggggatactctggagcgcgaatgacatgctcccaatttcttgaacgtgagcaaac 2073  
 DB 2318 CTTGGGGATCTCTGGAGCGCATGACATGCTCCAAAATTTCTTGAACAGTGAAGCAAC 2377  
 QY 2074 caaaaagccttctaagtagtagagcctgcactgctgcgtga 2115  
 DB 2378 CAAAAGCCTTCTTAATGAGTAGAGCCCTGCATGCTCCGTAA 2419

RESULT 8

AB000174

LOCUS AB000174 2604 bp mRNA

MAM

DEFINITION Porcine mRNA for endopeptidase 24.16, complete cds.

05-FEB-1999

AB000174  
VERSION AB000174.1 GI:1783131  
KEYWORDS endopeptidase 24.16 type M3; endopeptidase 24.16 (oligopeptidase M)  
SOURCE type M3.  
ORGANISM Sus scrofa Adult Liver cDNA to mRNA.  
REFERENCE Kato, A., Sugiyama, N., Saruta, Y., Hosokura, T., Yasue, H. and Hirose, S.  
AUTHORS Targeting of endopeptidase 24.16 to different subcellular  
TITLE compartments by alternative promoter usage  
JOURNAL J. Biol. Chem. 272 (24), 15313-15322 (1997)  
MEDLINE 97326108  
FEATURES  
source 1. .2604  
          /organism="Sus scrofa"  
          /db\_xref="taxon:9823"  
          /dev\_stage="Adult"  
          /tissue\_type="Liver"  
          1. .2604  
          /note="corresponding to exon3,5-16 of this gene;  
          endopeptidase 24.16 type 3"  
          116. .2161  
          /standard\_name="endopeptidase 24.16"  
          /note="oligopeptidase M: neurolysin :sBAP(soluble  
          angiotensin-binding protein) :MEP(microsomal  
          metalloendopeptidase)"  
          /codon\_start=1  
          /product="endopeptidase 24.16 type M3"  
          /protein\_id="BAA19066.1"  
          /db\_xref="GI:1783132"  
          /translation="MTLGRAMSPLOAMSSYVDGRNVRLWDLSPQIKRTEELIAQ  
          TKQYDDIGMDIEETVENCQLADLVEVYIVERTMLDFPQHVSDKEVRAASTBA  
          DKRLSDIFEMSMREDIFLRIVRLKTCDLGKPKPARYLEKSVKRGKRLHLPQ  
          VQNEIKAMKRSCLDFNKLNNEDDTFLVFSKAEGLALPDQFIDSLERDNDKIKYI  
          TLKYPFVPMKKCIPTERRKMEMAFNTRCKEENTIIQLLEPLRAKVLKLYSTH  
          ADFLVNTAKSTHHVFAFDLDSOKLPLGEAREPFIILNKKKECEKGFVDGKTN  
          AWDLHYMTQTEELKYSVDQELIKYEPFIEVVTGELLNIYQELGLSFEQVTDHVN  
          KSVLYIVKQKAGEVLQYFLDLIPREGRYNHACRGLQPGCLLPDGRMSVAALY  
          VNFQPRAGPSLLRHDEVRTPHFEGVHMQICAQDFARFSGTNVETDFVEVPSOM  
          LENWVWDTSLRLSKHYKDGSPITDLDLEKLVASRLVNTGLLTLRQIVLSKVDQSLH  
          TNTSLDAASEYAKYCTEILGVAATPGTNMPATFGLAGYDGOYVGYLWSEVFSMDMF  
          YSCFKKGIIMNPEVGMKRYRNLIKPGSLDMDMLQNLFLKRPKNQKAFKLSRGLHAP"  
BASE COUNT 820 a 478 c 594 g 712 t  
ORIGIN  
  
Query Match 86.4%; Score 1828.2; DB 4; Length 2604;  
Best Local Similarity 92.4%; Pred. No. 0;  
Matches 1920; Conservative 2; Mismatches 155; Indels 0; Gaps 0;  
  
Qy 39 cagagtggtgtccagagtttactcagatgacgttagaagagagagtgatgtctcc 98  
          |||||  
Db 85 CGGGTGTGGTTCGGGATTTTATTGAGAATGACTTTAGGAGAGAGCAATGTCCTC 144  
          |||||  
Qy 99 tcttcagcgaatgtcttctactatgtgtggtggcagaaatgttttaagatggatcttc 158  
          |||||  
Db 145 TCTTCAGCAATGCTTCTCTACTGTGGATGGCAGAAATGTTTAAAGATGGGATCTTTC 204  
          |||||  
Qy 159 accagagcaaatataaacaagactgagagtgatctatttgcagacaaacaggtgtacga 218  
          |||||  
Db 205 ACCAGGCAAAATTAAGAGAACTAGGAGCTGATTCGGCAGACCAACAGGTGTATGA 264  
          |||||  
Qy 219 tgcgttggaatgctcgtgtattgaggaagtaacttaacgaagactgtctgcaggcactggc 278  
          |||||



Db 1345 ACAGTTCTACTGGACCTCTATCAACAGGGAAGAAAATAACAACATGACGCCCTGCTTTGG 1404

Qy 1359 tctccagcctggctgctctctgctgctgagcagcgatgagcagtgctgctcctcgt 1418

Db 1405 TCTCCAGCCTGGCTGCTCTGCTGATGGAGCGGATGATGCTGCTGCTGCTGCTGCTGCT 1464

Qy 1419 ggtgaacctctcaagcagcagtgagcagctgctcctctctctctgagacacagcagtgag 1478

Db 1465 GGTGAACCTCTACAGCCAGCAGCAGGCTGCTCCCTCTCTCTGAGACACGATGAGGTGAG 1524

Qy 1479 gacttaattcatgagtttggtcacgctgagcatcagatattgtgcacagactgatttgc 1538

Db 1525 GACTTACTTCATGAATTTGGTCACGTCATGCTCATGATTTGTGCGCAGACTGACTTTGC 1584

Qy 1539 acgatttagcggacaactgtggaacctgactttgtagagtgccatcgcaaatgcttga 1598

Db 1585 ACGATTAGTGGACAATGTGNAACCTGACTTTGTAGAGGTGCCATCACAATGCTTTGA 1644

Qy 1599 aaattgggtggtggaactgacttccctcccgaaagtgttcaaaacattataaagatggaag 1658

Db 1645 AAACCTGGTGGACACTGATTCCTCCGAAGGCTGTCAAAACATTATAAAGATGGAAG 1704

Qy 1659 cctattgacagagatcgtgtgaaacctgtgtgttcttaggtgctgacacaggtct 1718

Db 1705 CCCTATTACAGATGATGCTGTGAAACCTTTGTGCTTGTAGACTGTGTCACACAGTCT 1764

Qy 1719 tctgacctgctgagcagattgtttgagcaagtgtgacgtctcttcatcaccacacatc 1778

Db 1765 CTTGACCTTAGCCAGATTTGTTTGACAAAGTTGATCAGTCTCTCACACACATTC 1824

Qy 1779 gctgagctgctgcaagtgaatatgccaatactgctgcagaatatattagagttgcagctac 1838

Db 1825 GCTGGATGCTCAAGTGAATGCCAATACTGCACAGAAATTTAGTGTGTCAGCTAC 1884

Qy 1839 tccaggcacaatatgcccagctaccttggacatttggcagggggatcacgatggccaata 1898

Db 1885 TCCAGGCACAAATATGCGACTATTTTGGACACTTGGCAGGGGGATGATGTCGCAATA 1944

Qy 1899 ttatggatatcttggagtgaaagtattttccatggatatattttacagctgttttaaaaa 1958

Db 1945 TTATGGATATCTTTGGAGTGAAGTCTTTCCATGGAGCATGTTTATAGCTGTTTAAAAA 2004

Qy 1959 agaaggataatgaatccagaggttggaatgaaatcacagaacctaactcctgaaacctgg 2018

Db 2005 AGAAGGCAATGAATCCAGAGGTGGAATGAATACAGAACTATCTCTGAAACCTGG 2064

Qy 2019 gggatctctggaaggatgacatgctcccaaatcttctgaaacgtgagccaaacaaaa 2078

Db 2065 GGGATCCTGGACGGCATGGACATGCTTCCAAAAATTTCTTGAACGTGAGCCAAACCAAAA 2124

Qy 2079 agcgttctaatgtagtagcctgcatgctcgtga 2115

Db 2125 AGCCTTCTATGATGAGAGCCCTGCTGCTCGTAA 2161

RESULT 9

LOCUS AB000175 2728 bp mRNA MAM 05-FEB-1999

DEFINITION Porcine mRNA for endopeptidase 24.16, complete cds.

ACCESSION AB000175

VERSION AB000175.1 GI:1783133

KEYWORDS endopeptidase 24.16 type M3.

SOURCE Sus scrofa Adult Liver cDNA to mRNA.

ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 2728)

Direct Submission

Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases.

Shigehisa Hirose, Tokyo Institute of Technology, Department of Biological Sciences; 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan (E-mail:shirole@bio.titech.ac.jp,

REFERENCE

AUTHORS Kato,A., Sugiura,N., Saruta,Y., Hosoiiri,T., Yasue,H. and Hirose,S.

TITLE Targeting of endopeptidase 24.16 to different subcellular compartments by alternative promoter usage

JOURNAL J. Biol. Chem. 272 (24), 15313-15322 (1997)

MEDLINE 97326108

FEATURES

source

Tel:045-924-5736, Fax:045-924-5824)

2 (sites)

Location/Qualifiers

1. .2728

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/dev\_stage="Adult"

/tissue\_type="Liver"

1. .2728

/notes="corresponding to exon3,4,5-16 of this gene;

mRNA

endopeptidase 24.16 type 3"

CDS

240. .2285

/standard\_name="endopeptidase 24.16"

/notes="oligopeptidase M :neurolysin :SBAP(soluble

angiotensin-binding protein) :MEP(microsomal

metalloendopeptidase)"

/codon\_start=1

/product="endopeptidase 24.16 type M3"

/protein\_id="BAA19067.1"

/db\_xref="GI:1783134"

/translation="MTIGREAMSPLOAMSSYTVDRNVLRWLSDPSOIKRRTEBELIAQ

TKOYDDIGMDIEEVTYENCLOADVEVKYIVERTMLDFPOHVSDDKVRRASTEA

DRLSRFDIEMSMREDIFLRVLKETCDLGKIKPEARVLEKSVKMGKRNGLHLPEQ

VQNEIKAMKRMSELICIDFNKNUNEDDTFLVSKAELGALPDDFIDISLEKTDNNKYI

TLKYPHYFPMVKCCIPENRRKMEAFNTRKEENTIIQEELPLRAKVAKLGLYSTH

ADPVELEMTAKSTHHVTAFLDLSOKLPLGEAREFIFLNKKKEEKEGFEDGKTN

AWDLHYMTQTEELKYSDQEIILKEYPFIEVVTGELLNIYOELLGLSFEOVTDHVMN

KSVTLATVKDKATGEVLGQFYLDLYPREGYNHAACFGLOPCLLPDGRSMVSAALV

VNFSQPRGRPSLLRHDEVNTYPERGHVHQICAOITDFARFSGTNVETDFVEVPSOM

LENWWDTSRLRLSKHYKDGSPITDILLKELVASRLVNTGLLTLOIOLVLSKVDQSLH

TNTSLDAASEYAKYTEILGVAATPGTNMPATFCHLAGGYDGYGYGLVSEVFSMDMF

YSCFKKEGIMNPEVMKRYNLILKPGSLDGMQLONFLKREPQNKAFLMSRLHAP"

BASE COUNT 847 a 518 c 616 g 747 t

ORIGIN

Query Match 86.4%; Score 1826.8; DB 4; Length 2728;

Best Local Similarity 92.3%; Pred. No. 0;

Matches 1921; Conservative 2; Mismatches 159; Indels 0; Gaps 0;

Qy 34 ctccgcagagttgggtgtccaggattttactcaagaatgcgttaggaagaagatgatg 93

Db 204 CACATCTGGGTGGTGGTCCGGGATTTTATTGAGAATGACTTTAGGAGAGAGAAGCAATG 263

Qy 94 tctcctctcaggaatgtcttctactgtgctgagcagaaatgttttaagatggat 153

Db 264 TCYCTCTTCAGGCAATGCTTCTTACTACTGTGGATGGCAGAAATGTTTAAAGATGGAT 323

Qy 154 ctttcaccagagcaaatataaacaagaactgagagctcatattgtgcagaccacaacaggtg 213

Db 324 CTTTCACCAGAGCAAAATCAAAAGAAAGAACTGAGGAGCTGATTGCGCAGACCAACAGGTG 383

Qy 214 taagatctgttgaatgctcgtgattgaggaagtaacttacgagaactgctcagagca 273

Db 384 TATGATGATATTGGAATGCTTGACATTTGAAGAGGTAACATTATGAGAACTGTTTTCAGGCA 443

Qy 274 ctggcagatgtagaagataagatatagtggaagagccatgctagactttccccagcat 333

Db 444 CTGGCAATGTGGAAGTGAAGTATATAGTGAACGACCATGCTAGACTTTCCCCAGCAT 503

Qy 334 gtatcctctgacaaagaagtacgagcagcagtagtacagaagcagacacaaagactttctcgt 393

Db 504 GTCTCCTCTGACAAAGAGTACGAGCAGCAAGTACAGAAGCAGACAAAAGACTTTCTCGT 563

Qy 394 ttgatattagatgagcatgagaggagagatatattttagagaagtgttcattacagaa 453

Db 564 TTTGATATTGAGATGAGCATGAGAGAAGATATCTTTCTGGAATTTGTCGTTTAAAGAG 623



FEATURES	Fax:	092-632-2742.	Location/Qualifiers	694 CCAGAGAAGTACAGAAATCAAATCAATCAAGAAAAAGATGAGTGAGCTCTGTATT 753
	source	1..3507		
CDS	/organism="Oryctolagus sp."	/db_xref="taxon:60556"	QY	601 gattttaacaaaaaacctcaatgagatgatactcttcttctgtattttccaaggctgaactt 660
	/tissue_type="liver"	154..2268		
	/codon_start=1	/product="endopeptidase"	QY	661 ggtgctcttcctgatgatttcattgacagtttagaaaaagacagatgatgacaagtataaa 720
	/protein_id="BAA02570.1"	/db_xref="GI:391849"		
	/translation="MIARCFSAVRGLHRVGRSLRFLKMTLGRVMSPLQAVSVYTAAG	RNVLRWLSPEQIKTRTEELIAOTKQVDSVGMIDIKDVTYENCLQALADVEVKYIVE	QY	721 attactttaaaataccacactatttccccctgcatgaagaataatgttgtatccccctgaacc 780
	RTMLDFPHQVSTGRVRAASTEADKRLSRFDIEMSMREDIFORVHLQETCDLEKIKP	EARYLEKSVKMRNGLHLPEEVQNEKSKMKRMSPLCIDFNKLNEDDTFLVFSKA		
	ELGALPDFIDSLKEMDDDKYKITLKYPHYPPVMKKCCIPETRRRMEAFNTRKEEN	TVIIQQLLPURAQVAKLGLYSTHADFLVLEMTAKTSRVTAFLLDDLSOKLPLRGEAR	QY	781 aqaagaagatggaatatgcttttaatacaaggtgcaaaagagagaaaaaccataattttg 840
	EFILSKKCEERKGFEDVGINAWDLHYMTQTEELKYSIDQEFINEIFFLEVTEG	LLNIYQELLSPEQVADAHVNPSTVLTVTKRATGEVLGVLDLYLPRGKNHIAA		
	CFGLQPGCLLPDGRMSLSVAALVNFQPVAGRPSLLRLRHDEVRVYFHEFGVHMOICA	QTFARSGTNVEDFVEVPSQMLENNWWDIDSLRLSKHYKDGNPIDADLLEKLIVAS	QY	934 AGAAGAGAGAAATGGAATGGCTTTTATACAAGATGCGAAAGAGGAAAAACCCGTCATCTG 993
	RLVNTGLTLRQIVLSKVQDSQSLHTNSLDAASEYARICTDILGVAATPGTNMPATFGH	LAGYDGGYGYGLMSEVPSMDMFYSCFKKEGIMNPEVGMKYRNLILRPGSLDGMML		
	QNFLOREPQKAFILMSRLQAP"	3412..3417	QY	901 gacttcgtcttgaatatgaactgcaaaagacacaaagccgcgttaacacagctttcttagat 960
	polyA_signal	3416..3421		
	polyA_signal	3420..3425	QY	1054 GACTTCGCTCTCGAGATGAACACTGCAAAAGAGCACAAAGCCGAGTAACAGCCTTTCTAGAC 1113
	BASE COUNT	939 a 794 c 874 g 900 t		
ORIGIN	Query Match	85.1%; Score 1799; DB 4; Length 3507;	QY	961 gatttaagccagaagttaaaaccccttgggtgaagcagaacgagattttatttgaatttg 1020
	Best Local Similarity	90.6%; Pred. No. 0;		
	Matches	1916; Conservative 2; Mismatches 197; Indels 0; Gaps 0;	QY	1114 GATTTAAGCCAGAAAATTAACACCGCTGGGTGAGGCGAAGACGAGAGTTCCTTTGAGTTTG 1173
	QY	1 atgatgccgggtgcttttgcgtgtgcgaagcctccgacaggttggtggtccaggatt 60	QY	1021 aagaaaaagaaatgcaaaagcaggggttttgaatatgatgggaaaaatcaatgctcgtggat 1080
	Db	154 ATGATCGCCCGGTGCTTTTCGGCTGTGCGAGCCCTCCACAGCGGTGGTGGTTCACAGGATT 213		
	QY	61 ttactcagaatgcgttagaagagaagtgcgtctctctcttccagcaaatgctcttccat 120	QY	1294 AAGGAGTACTTCCGATTGAGGTGCTACCGAGGGGCTGCTGAACATCTACCAAGGAGTTG 1353
	Db	214 CTATTCAAAATGACATTAGGAAGAGAAGTGATGTCACCTCTTCAGGCAGCTTTCTCCATT 273		
	QY	121 actgtggctgcagaaaaattttaagatggatctttccacagagacaaataaaacaaga 180	QY	1201 ttgggacttctcatttgaacaaatgacagatgctcattgtttggaaacaaagagtggttacactt 1260
	Db	274 ACTCAGCTGCGAGGAATGTTTTRAGATGGGATCTTTCACAGAGCAAAATCAAGACAAGA 333		
	QY	181 actgagagctcattgtgcagaccaaacaggtgtacgatcgtgttggaaatgctcggatt 240	QY	1354 CTGGGACTTTCAATTCAGCAAGTGGCCGACGCTCATGTTTGGAAACCCGAGTGTACACATT 1413
	Db	334 ACTCAGGAGCTCATTTGCCACACCAAGCAGGTGTATGATTTCTGTTGGAATGCTGGATATC 393		
	QY	241 gaggaagttaactacgagaactgctcgcagcactgagcagatgtagaagtaaaagtata 300	QY	1474 CCGAGGAAAGGAAAAATTAATCATGCAGCCTGCTTTGGTGTCTCCAGCCTGGCTGCTGCTG 1533
	Db	394 AAGGACGTGACTTACGAGAACTGCTTGACGGCGCTGGCAGATGTGGAAGTGAAGTACATA 453		
	QY	301 gtggaagagcactgctagactttcccgacatgtatcctctgacaaagaagtacagaca 360	QY	1381 cctgtggaagccgagatgatggcagtggtcctcctcgttgtaacttctcacagccagtg 1440
	Db	454 GTGGAAGAACCATGCTAGACTTTCTCAGCATGTTTCTACTGCAGAGAAAGTACGGGCA 513		
	QY	361 gcagatcacagaacagacaaagacttctcgttttgcattatgcagatgagcatgagaga 420	QY	1534 CCGGATGGGAGCGGATGCTGCTGTGGTGGCTCCCTCGTGGTGAACCTCTCCACGACGATG 1593
	Db	514 GCAAGTACAGAAGCGGACAGAGGGCTTTCTCGCTTTGATATTGATGAGCATGAGAGAA 573		
	QY	421 gatatattgagaaattgctcatttcagaaacactgactcgttggaagataaaacct 480	QY	1501 caggtgatgcacagatttggcacagactgatttgcagatttttagcgaaacaaatgtg 1560
	Db	574 GATATTTCAGAGAAATCGTTACATTACAGAAACCTGTGATCTGGAGAAAGATAAGCCG 633		
	QY	481 gaggcagacatacttggaaaagtccaattaaaaatgggaaaaagaaatgggctccatct 540	QY	1654 CATGTCATGCATCAGATTTGTGCACAGACTGATTTTGCACGATTTAGTGAACAAATGTG 1713
	Db	634 GAAGCCAGACCACTACTTGGAAAAGTCAGTTAAATTTGGAAGGAAGAAATGGGCTCCACCTT 693		
	QY	541 cctgaacaagtacagaatgaaatcaaatcaatgaagaaaaagaatgagtgagctatgtatt 600	QY	1621 tccctccgaagattgtcaaaaacattataaagattgggaagccctatttgcagacgactcgtt 1680
	Db			

Qy	1681	gaaaaacttcttgcctcaggctgggtcaacacacagggtctcttgaccctgagccagatgttt	1740
Db	1834	GAATAAATCTGTGCTTAGACATGGTCAACACACAGGTCTCTGACCTTCGCCAGATCGTT	1893
Qy	1741	ttgagcaaaattgatcagctctcttcataccaacacacacgctgagctgcaagtgaatat	1800
Db	1894	TTGAGCAAAAGTTGACCAAGTCTCTGCATACCAACTCATCCCTGGACGCTGCAAGTGAATAT	1953
Qy	1801	gccaaatactgctcagaaaatttaggagttgacgtactccaggcacaataatgccagct	1860
Db	1954	GCCAGGTACTGCACCGACATTTTAGGTGTTCGCGTACTCTCTGGACAACAACATGCCAGCT	2013
Qy	1861	acctttggacatttggcaggggatacagatggccaaatattatggatatctttggagtgaa	1920
Db	2014	ACTTTTGGCATTTGGCAGGGGATACGATGGCAGTATTATGGATATCTTTGGAGTGAA	2073
Qy	1921	gtatttccatgatatgtttttacagctgttttaaaaaaagggaataatgaatccagag	1980
Db	2074	GTGTTTTTCCATGGACATGTTTTTACAGCTGCTTTTAAAAAAGAGGAATAATGAATCCAGAG	2133
Qy	1981	gttgaatgaataacagaaacctaacttgaacacctgaaacctggggatctctggacggcatggac	2040
Db	2134	GTGGAATGAAATACAGAAACCTTAATCTGAGACCTGGGGGATCTCTGGACGGCATGGAC	2193
Qy	2041	atgtccacaatttcttgaacgtgagccaaacccaaagcgttccttaagttagtaggac	2100
Db	2194	ATGCTCCAAAATTTCTTGCAACGTGAGCCAAACACGAAACCGTTCTCTGAATGAGCCGGGC	2253
Qy	2101	ctgcatgctcgtga 2115	
Db	2254	CTGCAGGCTCCCTGA 2268	
RESULT	11		
RNRNAME		RNRNAME	2448 bp mRNA
LOCUS		R. norvegicus	mouse endopeptidase.
DEFINITION		R. norvegicus	mouse endopeptidase.
ACCESSION		X87157	
VERSION		X87157.1	GI:987086
KEYWORDS		metalloendopeptidase; neurolysin; neurotensin-degrading neutral protein.	
SOURCE		Norway rat.	
ORGANISM		Rattus norvegicus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
		Rattus.	
REFERENCE		1 (bases 1 to 2448)	
AUTHORS		Dauch, P., Vincent, J.P., and Checler, F.	
TITLE		Molecular cloning and expression of rat brain endopeptidase	
JOURNAL		J. Biol. Chem. 270 (45), 27266-27271	(1995)
MEDLINE		96070836	
REFERENCE		2 (bases 1 to 2448)	
AUTHORS		Checler, F.	
TITLE		Direct Submission	
JOURNAL		Submitted (01-MAY-1995) F. Checler, Inst. de Pharmacologie	
		Moléculaire, et Cellulaire, Centre National de la Recherche	
		Scient., UPR 411, 660 Route des Lucioles, Sophia Antipolis, 06560	
		Valbonne, FRANCE	
COMMENT		Overlaps with D13310.	
FEATURES		Location/Qualifiers	
source		1..2448	
		/organism="Rattus norvegicus"	
		/strain="Sprague Dawley"	
		/db_xref="taxon:10115"	
		/clone_lib="ZAP II"	
		/clone="pBluescript"	
		144..2258	
CDS		/EC_number="3.4.24.16"	
		/note="neurolysin"	
		/codon_start=1	
		/product="neurotensin-degrading neutral	

```

metalloendopeptidase"
/protein_id="CAA60630.1"
/db_xref="GI:987087"
/db_xref="SWISS-PROT:P42676"
/translation="MITTCLSTLGRHAGGSRSLQTLMTLGGKELASPLQAMSSYTAAG
RNYRLWDLPSEQIETRTQALQIQTQVOTVGTIALKEVTYENCVLQADIEVTYIVE
RWMLDPQHVSSDREVRAASTEADKKLSFEDIMSMREDVFQIRVHLOETDLEKTKP
EARRYLEKSTIKMGKRNGLHLSEHIREIKSMKKRMSELCIDFNKLNENDDTSLVFSKA
ELGALPPDFIDSLSEKTDDEQIKVLYIKAFVYPMKKKCCVPETPRKMEMAFHPRCKQEN
TALQLQLPLRAQVAKLLGYNTHADEFLNTAFKTSRVAALDDLUSSKDKPLGEAER
EFTLSLKKCEBERGEYFDGKINAWLHYITQTTELKYSPQESLKEFYETVEVTEG
LLSITVQLLEGLSEFQYDPADHWNKSVSLYTKDKATGEVLGQFYLDLYPREGKYNHAA
CFGLPGCLLPDGRSMMSVAALVNFNSQVAGRPSSLRHDEVRTYFHEFGYVHMQLCA
QTDFARPSTNVTEDTSEVPSOMENWMDVDSLKLSKHYYKDGPHTIDELLEKLVAS
RLNTGLLLTQVLSLQDLSLQATLNTATLDAASYAKYCTEILGVAAVTPGTNPATFGH
LAGGYDQGYGVLMSEVFE SMDMFSCFKKEGINNPEVGMKIRNLLIKPGSLDGDMDL
QNFLQREPNQKAFILMSRGLNGS"
683 a 566 c 622 g 577 t

```

Query Match 78.28; Score 1653.4; DB 10; Length 2448;

```
Query Match 78.28; Score 1653.4; DB 10; Length 2448;
```

		Best Local Similarity	86.3%	Pred. No. 0;	Mismatches	288;	Indels	0;	Gaps
		Matches 1825;	Conservative	2;					
Qy	1	atgatgcgccgggtgccccttttggcgtgcgaagccctccgacagtgttggttcaggatt	60						
Db	144	ATGATCACCCCTGTGCCTTTCGACACTCTGGAGGCCCTCCACAGAGCTGTGGGTCCACGGCTT	201						
Qy	61	ttaactcaaatgacgttaggaagaagatgatgtctctcttcaggccaatgctcttcccat	120						
Db	204	CAGCTTGACAATGACACTGGGAAGAAGAGCTGGCTTCTCCTTTCAAGCATGCTCTTCCTAC	261						

[illegible]

384 AAAGAAGTAACTTATGAGAACTGCTCGAGTGTGGTGTACATAGAAGTGACCTACATA 44  
Db  
301 gtgaaaggaccatgctagactttcccccagcatgtatcctctgacaaagaagtaacgagca 36  
Qy

50	CTGGAAGACCATCTGGACTTCCCTCAGCAGCGTGTCTATCTGACAGAGAAGTGGCGGCT	50
444		444
Db		Qy
361	gcaagtacagaagcagacacaaagacttctcgttttgaatattgagatgagcatgagaga	422
504	GCACGACAGAGAAGCAGACAAAAGGTTTCTCGTTTTGTATTGAGATGAGCATGAGAGA	566
Db		Db

Qy	421	gatatattgagaataattgttcatttacagaaacctgtgatctggggaagataaacct	480
Db	564	GATGTGTTTCAGAGAATTGTTTCATTTTACAGAACATCGGATTTGGAGAAATAAAGCT	621

[illegible]

Db	694	TCAGACACATAGAGAAATCAATCATCATGAGAGAGAGAGATGAGCTATGCAATT	74
Qy	601	gattttaacaaaacctcaatgaggatgatcacttccttcttatttccaaagctgaactt	666
Db	744	GACATTCAACAGAACCTCAATGAGAGACACTTCCCTTGATTCTTCAAGGCTGAACCT	800

Qy	661	gggtctctctgatgattcattgacaggttataaagacagatgatgacagaatataa	72
	804	GGTGTCTTCTGTATGACTTCATTGACAGTTTAGAAAAACAGATGAAGACAGTACAA	86

Qy	721	attacctaaatatccacactatttccctgtcatgaagaatgttgtatccctgaacc	780
Db	864	GTTTACCTTAAAGTATCCTCCTACTACTTCCCTGTCTATGAAGAAATGCTGTCCCTGAAACT	923
Qy	781	agaagaagatgaaatggcttttaatacaaggtgcaaaagaggaaaaacacataatttgg	840
Db	924	CGAAGGAAGATGAATAGGCTTTTCATACAGGTGCAAAACAGGAAAAACACCGCAATTTCGTG	983
Qy	841	cagcagctactcccactgcgaaccaaggtgggccaactactcgggtatagcacacatgct	900
Db	984	CAACAGCTCTCCCGCTGCGAGCTCAAGTGGCCCAAGCTGCTGGGCTATACACCCATGCT	1043
Qy	901	gacttgccttgaatgaacactgcgaagagcacaagcgcgtataacagcyyttctagat	960
Db	1044	GACTTTGCTCTTGAACATGAACATCGCGAAGAGCACACGCCGCGTGGCGCTTTCTAGAT	1103
Qy	961	gatttaagccagaagttaaacccttgggtgaagcagaacagagatttattgaatttg	1020
Db	1104	GATTTAAGCCAGAAATGAAACCCCTTGGGTGAGGCAGAACGGGAGTTTATTTTGAGCGTG	1163
Qy	1021	aagaaaaaggaatgcgaagacagaggggttttgaatatgatggaaaaatcaatgcctgggat	1080
Db	1164	AAGAAGAAGGAATGTCAAGAGCGAGGTTTTTGAATACGACGCGGAAGATCAATGCATGGGAC	1223
Qy	1081	ctatactactaactgactcagacagaggaactcaagtattccatagaccgaagttccctc	1140
Db	1224	CTCCTACTTACTATGACACAGACAGAGGAGCTCAAGTACTCGGTGGACACAGGATCCCTT	1283
Qy	1141	aaggaataacttcccaattgaagtgatcactgaaggttgctgaacctaccaggaattg	1200
Db	1284	ARGAGTACTTCCCCANTGAGTGGTGCAGGAAGGCTTGTCTAAGCATCTACAGGAGCTG	1343
Qy	1201	ttgggactttcatttgaacaaatgcagatgcctcatgttttggaaacaagagtgtcacactt	1260
Db	1344	CTGGGACTTTTCATTTGAACAAGTCCCGATGCTCATGTTTGGATATAGAGCGTTTCATT	1403
Qy	1261	tatactgtgaaggataaagctacagagagaagbattgggacaggttctctatttggacctctat	1320
Db	1404	TACACTGTGAAGGATTAAGCTACAGGAGAGTTCCTTGGGGCAGTGTCTACCTGGACCTCAT	1463
Qy	1321	ccaaggaaggaataacaaatcatcgccctgtcttcggtctccagcctggctgcctcttg	1380
Db	1464	CCAAGGAAGGAATAATACAACTCATGAGCCTGCTTTCGGTCTCCAGCCAGGCTGGCTTCTC	1523
Qy	1381	cttgatggaagccggatgatggcagtgggctgcccctcgttggtagaacttctcacagccagtg	1440
Db	1524	CCTGATGGCAGTCCGATGATGCTGTGGCGCCCTGGTGGTCAACTTCTCTCAGCCGTGA	1583
Qy	1441	gcaggtgtccctctcctctctctgacacgacgaggtgaggacttacttctatgagttcgtt	1500
Db	1584	GCAGTGGGCCCTCTCTCTGTGAGCATGATGAAGTTCGGGACTTACTTCCACAGATTCCGT	1643
Qy	1501	cacgtgatcatcagatttgtcacagactgattttgcacgatittagcggaacaaatgtg	1560
Db	1644	CACGTATGTCATCAGATCTGTGGCAGACTGACTTTGCACGATTCAGTGGGAACAAACGTG	1703
Qy	1561	gaaactgacttttagaggtgcccactgcaaaatgcttgaaaaattgggtgtgggagctcgat	1620
Db	1704	GAACATGACTTTGTAGAGGTGCCATCACAATGTCTTGAAAACCTGGGTGGGACGTTGAT	1763
Qy	1621	tcctccgaagattgtcacaacattataaagatggaagccctatttgacagcgtatctgctt	1680
Db	1764	TCCCTGGGAAACTGTCAAAGCATTAAGAATGGACACACCCTATCACAGACGAGCTGCTC	1823
Qy	1681	gaaaaacttgttctctctaggtggttcaacacaggtgtcttctgaaccttgccacagattggt	1740
Db	1824	GAGAAGCTTGTGTCTTCTAGGTGGTTCACACAGGCTTCTGAGCGTCCGCGCAAAATGTT	1883
Qy	1741	ttgagcaaaattgatcagctctcttcatacaacacatcgctggatgctgcaagtgaatat	1800
Db	1884	TTGAGCAAAAGTTGACCAATCTCTCCATACCAATGCGACTCTGGGATGCGCGAGGGAATAC	1943
Qy	1801	gccaaatactgctcagaaatatattagagttgcagctactctccaggcacaataatctcagact	1860

[illegible]

```
/codon_start=1
/product="Unknown (protein for MGC:3978)"
/protein_id="AAH04985.1"
/db_xref="GI:13436419"
/translation="MIRACLLAVSLRVRGSRILLRMTLGRVNSPLQAMSSYTVAG
RNVRLDLSPOIKITRTEELIVQKVYDAVMGLIEEIVYENCLOALADYEVKIVE
RTMLDPPOHVSDDREVAASAEADKRLSRPDIEMSRGDIIFERIVHLEQCDLQKIRP
EARVLEKSIKMGKRNGLHLPQVONETKSMKRMSELCIDFNKLNEDDTFLVFSKA
ELGALPHDFIDSLKTDKDKYIKLKYPHYFVPMKKCIPETRRMEMAFNTRKEEN
TIILQOLPLRKTAKVALLGYSHTADFVLEMTAKSTSRVTAFLDDLQSLKPLGEAER
EFILNKKKECKDFEYDGRINAWDLYYMTQTEELKSIDQFLKYEYFIEVYTEG
LLNTYELLGLSFQMDTAHWNKSVTLTVKDRATGEVLQGFYLDLYPREGKNHAA
CFNGLOQLCLPDGSRMVAALVFNOPVAGRPSLLRHDEVRYEFHEFGVHMQICA
QVSFFPVPNLPISEFKLLDCCOVSAPNRFQKLVNFKNFGLLQNLQNFPPVLFHN
IFLGGSTSERCLNLSLWKNKTKICKERLQILCFLLLL"
BASE COUNT 790 a 522 c 592 g 733 t
ORIGIN

Query Match 72.2%; Score 1526; DB 9; Length 2637;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1529; Conservative 1; Mismatches 6; Indels 0; Gaps 0:

QY 1 atgacgcgcgcgtgctcttggctgtggaagcctccgcagagttgggtgtccaggatt 60
DB 117 atgacgcgcgcgtgctcttggctgtggaagcctccgcagagttgggtgtccaggatt 176

QY 61 ttactcagaatgacctaggagaagaagtgtgtctctctctcaggaatgtctctccat 120
DB 177 ttactcagaatgacctaggagaagaagtgtgtctctctcaggaatgtctctccat 236

QY 121 actgtgctgcgcagaaatgttttaagatggatctttcaccagacaaattaaacaaga 180
DB 237 actgtgctgcgcagaaatgttttaagatggatctttcaccagacaaattaaacaaga 296

QY 181 actgaggagctcattgtgcagaccacacaggtgtacagatgctgttggaaatgctcggtatt 240
DB 297 actgaggagctcattgtgcagaccacacaggtgtacagatgctgttggaaatgctcggtatt 356

QY 241 gaggaagtaacttacagaactgtctgcagcactcggcagatgtagaagtaaaagtata 300
DB 357 gaggaagtaacttacagaactgtctgcagcactcggcagatgtagaagtaaaagtata 416

QY 301 gtggaagaccatgactgacttcccagcatgtatcctctgacaaagaagtcagaca 360
DB 417 gtggaagaccatgactgacttcccagcatgtatcctctgacaaagaagtcagaca 476

QY 361 gcaagtacagaagcagacaaagactttctcgttttggatttggatgagcatgagagga 420
DB 477 gcaagtacagaagcagacaaagactttctcgttttggatttggatgagcatgagagga 536

QY 421 gatattattgagagaattgttcattacagagaacctgtgactcgtgggaagataaaacct 480
DB 537 gatattattgagagaattgttcattacagagaacctgtgactcgtgggaagataaaacct 596

QY 481 gaggcacagacatcttgaaaagtcaatataatgggaaaagaatgggctccatctt 540
DB 597 gaggcacacagatcttgaaaagtcaatataatgggaaaagaatgggctccatctt 656

QY 541 cctgacaagtcagagaatgaatcaaatcaatgaagaaaagaatgagtgagctatgtatt 600
DB 657 cctgacaagtcagagaatgaatcaaatcaatgaagaaaagaatgagtgagctatgtatt 716

QY 601 gatatttaaaaaaacctcaatgagatgataccctcctctgtattttccaaaggtgaactt 660
DB 717 gatatttaaaaaaacctcaatgagatgataccctcctctgtattttccaaaggtgaactt 776

QY 661 ggtgctctcctgatgtattcattgacagtttagaagaagacagatgatgacaagtataa 720
DB 777 ggtgctctcctgatgtattcattgacagtttagaagaagacagatgatgacaagtataa 836

QY 721 attaccttaaatatccacacatttccctgtcatgaagaatgttgcctccagaaacc 780
```

```
DB 837 ATTACCTTAAATAATCCACACTATTTCCTGTGTCATGAAGAAATGTTGTATCCCTGAAACC 896
QY 781 aqaagaagatggaatggcttttaatacaaggtcaaaagagagaaacacacataatttg 840
DB 897 AGAAGAAGGATGGAATGGCTTTTAAACAGGTGCAAGAGGAAGAAACCAATAATTGG 956
QY 841 caagagctactccccactgcgaaccagggtggcgaactactcgggttagacacacatgct 900
DB 957 CAGCAGCTACTCCCACTGGGAACCAAGGTGGCGCAAACTACTCGTGTATAGCACACATGT 1016
QY 901 gacttcgctcttgaaatgaacactgcaagagacacaaagcgcgcgttaacagcyttttctagat 960
DB 1017 GACTTCGTCCTTGAATGAACACTGCAAGAGACACAAAGCCGCTAACAGCCCTTCTTAGAT 1076
QY 961 gatttaagccagaagttaaaaccccttgggtgaagcagaagagagaggtttattttgaatttg 1020
DB 1077 GATTTAAGCCAGAAAGTTAAACCCCTTGGGTGAAGCAGACGAGAGTATTTATTTGAATTGG 1136
QY 1021 aagaaaaaggaatgcaagacagagggtttttgaatatgatgggaaaaatcaatgcctgggat 1080
DB 1137 AAGAAAAAGGAATGCAAGACAGAGAGGGTTTTGAATATGATGGGAAAAATCAATGCCCTGGGAT 1196
QY 1081 ctatatctactacatgactcagacagacagaaactcaagtagttccatagacacaaagagttctctc 1140
DB 1197 CTATATTACTACATGACTCAGACAGAGAACTCAAGTATTCATATAGACCAAGAGTTCCTC 1256
QY 1141 aaggaatacttcccgaattgaggtggtcactgaagccttgctgaacacacctaccagaggttg 1200
DB 1257 AAGGAATACTTCCCAATTGAGGTGCTCACTGAAGGCTTGTGCAACACCTTACCAGGAGTTG 1316
QY 1201 ttgggactcttcatttgaacaaatgacagatgctcatgttttggaaagaagtggttacactt 1260
DB 1317 TTGGGACTTTCATTTTGACAAATGACATGCTCATGTTTGGAAACAAGTGTTACACTT 1376
QY 1261 tatacttgaaggataaaagctacaggaagaagtatttggacagttctctatttggacctctat 1320
DB 1377 TATACTGTGAAGATAAAGCTACAGGAGAAAGTATTTGGGACAGTTCTATTTGGACCTCTAT 1436
QY 1321 ccaaggggaagaaaaatacaatcatcgtgcctcgtcgttcctccagcctgctgcctcttg 1380
DB 1437 CCAAGGGAAGGAAAAATACAATCATCGGCCCTGCTTCGGGTCTCCAGCCTGGGCTGGCTTCG 1496
QY 1381 cctgtatgaagcggatgatggcagtgctgcctcgtcgttgtaacttctcacagccagtg 1440
DB 1497 CCTGATGAAGCCGGATGATGCGCAGTGGCTGCCCTGCTGGTGAACCTTTCACAGCCAGTG 1556
QY 1441 gaagtcctcctctctcctgagacacagcagaggtgaggaacttaacttgaattggttggat 1500
DB 1557 GCAGTCTGCTCCTCTCTCTGAGACACGACGAGGTGAGGACTTACTTTTCATGAGTTGGT 1616
QY 1501 cagtgatgcatcagatttggcacagactgatttt 1536
DB 1617 CACGTGATGTCATCAGATTGTGTCACAGAGGTGAGTTT 1652
```

```
RESULT 13
BC001644
LOCUS BC001644 2646 bp mRNA PRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to thimet oligopeptidase 1, clone MGC:2361
IMAGE:2958628, mRNA, complete cds.
ACCESSION BC001644
VERSION BC001644.1 GI:12804472
KEYWORDS MGC..
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2646)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```



USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadan@systemsbio.org](mailto:amadan@systemsbio.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 2 Row: 0 Column: 6...

FEATURES

Location/Qualifiers  
1..2646  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:2361 IMAGE:2958628"  
/tissue\_type="Kidney, renal cell adenocarcinoma"  
/clone\_lib="NIH\_MGC\_14"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
126..1949  
/codon\_start=1  
/product="Similar to thimet oligopeptidase 1"  
/protein\_id="AAH01644.1"  
/db\_xref="GI:12804473"  
/translation="MIARCLLAVSLRRVGGSRILLRMTPLGRVMSPLQAMSSYTVAG  
RNLWDLSPGDIKTRTEELIVOTKQVDAVGLMTEEVYENCILQALADVVKVIVE  
RMLDFPOHVSDDKVRASAEADKRLSPEDLEMSMRGDIFFERIVHLQETCDLGIKP  
EARRLEFSIKMGKNGLHLPQVONEIKSMKRMSELCIDENKNLNEDDTFLVFSKA  
EUGALPHDFIDSLKDDDDKDKITIKLIPHYPMKKSCCIPETRRRMENAFNRCKEEN  
TIILOQLPLKRLGKLYSTHADFLVEMNTAKSTSRVAFDLDSQKLPGLBAER  
ETILKKKECKDFEYDGIKINAWDLVYMTQTEELKYSIDQEFLEKFFPEVTEG  
LNTLOEILGLSFEQMDAHVWKNVLTLYTKDKATGEVLGOFYLDLYPREKYNHAA  
CFGLQPLGLPDSRMVAALVNFSPQVAGRPSLLRHDEVRTYFHEGHVWHQICA  
QVSFFPPVNLPISEFKLLDCCQVSAPNRFQKLVNFKLVFLGLLQNFPIPVLFHN  
IFLLGGSSYSEKLLNLSLWKNKTKICKERLIQILCFLLIL"

CDS

BASE COUNT	792 a	524 c	597 g	733 t
ORIGIN				
Query Match	72.2%; Score 1526; DB 9; Length 2646;			
Best Local Similarity	99.5%; Pred. No. 0;			
Matches 1529; Conservative	1; Mismatches	6; Indels	0; Gaps	0;
QY	1	atgacgcccgtgcttttggctgtgcgaagcctccgcagagttggtggtccaggatt	60	
Db	126	ATGATCGCCGGTCCCTTTGGCTGTGCGAAGCCTCCGACAGTTGGTGTTCAGGATT	185	
QY	61	ttactcagaatgacgttaggaagaagtgaatgtctctcttcttcaggcaatgtcttcttat	120	
Db	186	TTACTCAGATGACGTTAGGAAGAAGTAGTGCTCTCTTCACGCAATGCTTCCTAT	245	
QY	121	actgtggctggcagaaatgttttaagatgggactcttcaccagagcaaatataaacaaga	180	
Db	246	ACTGTGCTGGCAGAAATGTTTAAAGATGGATCTTTACCACAGACAAATTAACAAAGA	305	
QY	181	actgagagctcatgttgcagaccacaacagggtgtacgatgctgttgaaatgctcggtatt	240	
Db	306	ACTGAGGAGCTCATTTGTGCAGACCAACAGGTGTACGATGCTGTTGGAATGCTCGGTATT	365	
QY	241	gaggaagtaacttacgagaactgtctgcaggcactggcagatgtagaagtaaaagtatata	300	
Db	366	GAGGAAGTAACTTACGAGAACTGTCTGACAGCACTGGCAGATGTAGAAGTAAAGTATATA	425	
QY	301	gtggaaggaacctgctagactttccccagcatgtatcctctgacaaagaagtacgagca	360	
Db	426	GTGGAAGGACCATGCTAGACTTTCCCCAGCATGTATCCTCTGACAAAGAGTACGAGCA	485	

QY	361	gcaagtacaga	gcagacaaag	actttctcgt	tttgatattg	agatgagcag	gagga 420
Db	486	GCAAGTACAG	AACACAGAA	AGACTTTCTC	GTTTTGATAT	TGATGAGATG	AGAGGA 545
QY	421	gatatattg	agagaattgt	tcatttacag	gaacacctgt	gatctgggga	agataaaacct 480
Db	546	GATATATTT	GAGAGAAAT	TGTTCAATTT	TACAGAAACCT	TGTGATCGG	GAAGATAAAACCT 605
QY	481	gaagccagac	gatacttgaa	aaagtcaatt	aaatgggaaa	agaaatggc	ctccatctt 540
Db	606	GAGCCAGAG	TACTTGGAAA	AGTCAATTA	AAATGGGAAA	AGAAATGGG	CTCCATCTT 665
QY	541	cctgaacaag	tacagaaatg	aaatcaaatc	aatgaagaa	gaatgagtg	agctatgtatt 600
Db	666	CTTGAAAC	AGTACAGAA	TGAATCAAT	CAATCAAG	AAAGATGAG	TGAGCTATGTATT 725
QY	601	gattttacaa	aaacacctca	atgaggatg	atcctctctg	tattttccaa	ggctgaaacctt 660
Db	726	GATTTTAA	CAAAAACCT	CAATGAGAT	GATACCTTC	CTTGTATTTT	CCAAGGCTGAACCTT 785
QY	661	ggtgctcttc	ctgatgaatt	ttcattgac	gttttagaaa	agacagatg	atgacaaagtataaa 720
Db	786	GGTGCCTT	CCTCATGAT	TTTCATTG	ACAGTTTAG	AAAAAGAC	AGATGACAAAGTATAAA 845
QY	721	attaccttaa	aatatccaca	ctatttcc	ctgtcatg	aaagaaatg	tgttattccctgaaacc 780
Db	846	ATTACCTT	AAAAATAT	CCACACTAT	TTCCCTGT	CATGAAGA	AAATGTTGTATCCCTGAAACC 905
QY	781	agaagaag	gatgaaatg	gcttttaata	cagaagtg	caaaagaga	aaaaaacacataatttg 840
Db	906	AGAAGAAG	GATGAAAT	GGCTTTTAA	TACAAGTG	GCAAGAGG	AAAAACACCAATAATTTTG 965
QY	841	cagcagctac	ctccactcga	aaccaagg	tggccaaac	tactcgtt	atagcacacatgct 900
Db	966	CACAGCTAC	TCCACTGCG	AACCAAGG	TGGCCAAAC	TACTCGGT	TATACACACATGCT 1025
QY	901	gacttcgtc	cttgaatga	acactgcaa	agacacag	cgcgttaac	agcagcytttctagat 960
Db	1026	GACTTCGT	CTCTGAAT	GAACACTG	CANAGAGC	ACAGCCGCT	TAACAGCCTTCTAGAT 1085
QY	961	gatttaagc	cagaagtt	aaaaacct	ttgggtgaa	gcagagag	cagagatttatttgaattg 1020
Db	1086	GATTTAAG	CCAGAA	GTAAAA	CCCTTGGG	TGAAGC	AGACAGAGTATTATTGGAATTTG 1145
QY	1021	aegaaaaag	aatgc	aaagacag	gggtttttg	aatatgat	tgaggaaaaatcaatgcctggat 1080
Db	1146	AAGAAAAAG	CAATGCA	AAAGAC	AGGGGTTT	TGAATATG	ATGATGGAAATCAATGCTGGGAT 1205
QY	1081	ctatactac	atgactgac	agaggaag	aaactcaag	tattccat	agaccagagttcctc 1140
Db	1206	CTATATTAC	TACATGACT	CAACAGAG	GAAGAACT	CAAGTATTC	CAATAGACCAGAGTTCCTC 1265
QY	1141	aaggaatac	ttcccaattg	aggttggtc	actgaagg	cttgc	tgaaacacctaccaggagttg 1200
Db	1266	AAGSAATAC	TTCCCAAT	TGAGGTGG	TCTCACTGA	AGGCTTGT	GTAACACCTACCAGGAGTTG 1325
QY	1201	ttgggacttc	atttga	aaaaatg	acagatg	ctatgtt	tggaacaagagtggttacactt 1260
Db	1326	TTGGGACT	TTTCATTT	GAACAAAT	GACAGATG	CTCATGCT	TTTGGAAACAAGAGTGTGTACACTT 1385
QY	1261	tatactgaa	gataaaag	ctacagga	agattggg	acagttct	tatttggaccttat 1320
Db	1386	TATACGTG	TGAAGGAT	AAAGCTAC	AGGAGAA	GTATTTGG	GCAGTCTTAT 1445
QY	1321	ccaaggga	agaaaaat	acacatc	atg	cggtctc	ccagcctggctgcctctg 1380
Db	1446	CCAAGGAG	GAAGAA	TACATCAT	TCCGCCCT	GTCTCGG	TCTCCAGCCTGGCTGCCCTCTG 1505
QY	1381	cctgatagg	aagccggat	gatg	ggcagtg	ggcctc	ogtggtaacttctcacagccagtg 1440
Db	1506	CCTGATGG	AAAGCCG	GGATGATG	GGCAGTGG	CTCGCT	CTCGTGTGAACCTTCTCACAGCCAGTG 1565
QY	1441	gcaggctgc	tctctcct	ctgagaca	gcagaggt	gagga	cttacttcatgagtttggt 1500

Db 1566 GCAGGTGCTCCCTCTCCTGAGACACGACGAGGTGAGGACTTACTTTCATGAGTTGGT 1625  
Qy 1501 cacgtgatcagcagatttgcacagactgatttt 1536  
Db 1626 CACGTGATGCATCAGATTGTGACACAGGTGAGTTT 1661  
RESULT 14  
AB033052 5343 bp mRNA PRI 11-NOV-1999  
DEFINITION Homo sapiens mRNA for KIAA1226 protein, partial cds.  
ACCESSION AB033052  
VERSION AB033052.1 GI:6330675  
KEYWORDS  
SOURCE 'Homo sapiens brain cdna to mRNA, clone\_lib:pBluescriptII SK plus  
clone:fh04414'.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (sites)  
Nagase,T., Ishikawa,K., Kikuno,R., Hirose,M., Nomura,N. and  
Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes. XV.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro.  
JOURNAL DNA Res. 6 (5), 337-345 (1999)  
MEDLINE 20039619  
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-1999) to the DDBJ/EMBL/GenBank databases. Osamu  
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail:cdnainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/,  
Tel:++81-438-52-3913, Fax:++81-438-52-3914)  
FEATURES  
source  
1. 5343  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="pbluescriptII SK plus"  
/tissue\_type="brain"  
1. 1345  
/gene="KIAA1226"  
1. 1345  
/gene="KIAA1226"  
/note="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1226 protein"  
/protein\_id="BAA86540.1"  
/db\_xref="GI:6330676"  
/translation="PETERREWAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHAD  
FVLEMTAKSTSRVTAFLDLSQLKPLGEAREFILNLKKKECKDRGPEYDKINAW  
DLIYMTQELKSIDOELEKPEFYEVVTEGLNTYQELGLSPQMTDAHWNKS  
VLYTVKDKATGVLGYFDLPLREGKYNHAAFCFLQPGCLLPDGSRMMAVALVYN  
FSQPVAGRSLLRHDEVRTYFHEFGVHMQICATDFARESGTNVDFDFEVSQMLE  
NWYVDLSRLKSHYKDGSPIDADLEKLVASRLVNTGLTLROI VLSKVDOSLHTN  
TSLDAASEYAKYCEILGVAATPCTNNPATFGLHAGYDGYGYLWSEVFSMDMEYS  
CFKREGIMNFEVGMKYRNILKPKGSLDGMHLNLFKRPENQKAFMLSRGLHAP"  
BASE COUNT 1607 a 1060 c 1107 g 1369 t  
ORIGIN  
Query Match 63.68; Score 1344.2; DB 9; Length 5343;  
Best Local Similarity 99.99; Pred. No. 0;  
Matches 1343; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 771 cccctgaaccgaagaagatggaattggcttttaatacaagggtgcaaaaggaggaacac 830  
Db 1 cccctgaaccgaagaagatggaattggcttttaatacaagggtgcaaaaggaggaacac 60  
Qy 831 cataatttgcagcagctactcccactgcgaaccaagggtggccaactactcgttatag 890

Db 61 CATAATTTTCAGCAGCTACTCCCACTGCGAACCAAGGTGGCCAAACTACTCGGTATAG 120  
Qy 891 cacacatgctgaacttcgtcccttgaaatgaacactgcgaagagcaaaagccggttaacagc 950  
Db 121 CACACATGCTGACTTCGTCCTTGAATTAACACTCAAGAGAGCAAGCCGCGTAACAGC 180  
Qy 951 yttctagatgatttaagccagaagttaaaccccttgggtgaagcagacagagagtttat 1010  
Db 181 CTTTCTAGATGATTAAGCCAGAGTTAAACCCCTTGGGTGAGCAGACAGAGCTTAT 240  
Qy 1011 ttgaaatttgaagaaaaaggatgcaaaagacaggggtttttgaatatgatggaaaaatcaa 1070  
Db 241 TTTGAATTTGAAGAAAAAGGAATGCAAGACAGAGGGGTTTGAATATGATGGAAAAATCAA 300  
Qy 1071 tgcctgggagtctattactactacagactcagacagagagaactcaagtattccatagacca 1130  
Db 301 TGCCTGGGACTATATATTACTACTGACTCAGACAGAGGAATCAAGTATTCATAGACCA 360  
Qy 1131 agagtccctcaaggaatacttcccgaattgagtggtgcactgaaggtcttgcaaacaccta 1190  
Db 361 AGAGTTCTCTCAAGGAATACTTCCCAATTTGAGGTGCTCACTGAAGGCTTGTGACACCTA 420  
Qy 1191 ccaggagtgttgaggacttctcatttgaaacaaatgacagatgctcattgttggaaacaag 1250  
Db 421 CCAGGAGTTGTTGGGACTTTCATTTGAACAAATGACAGATGCTCATGTTTGGAAACAAG 480  
Qy 1251 tgttacactttactactgtaagataaaagctacagagaagattattggacagttctatt 1310  
Db 481 TGTACACTTTATCTGTGAAGGATAAAGCTACAGAGAAGTATTGGACAGACTTCTATTT 540  
Qy 1311 ggacctctatccaaaggggaagaaatacaatcatcgcgcctgcttcgcctccagcctgg 1370  
Db 541 GGACCTCTATCCAAAGGGAAGAAAAATCAATCATCGCGCCTGCTTCGCTCTCCAGCCTGG 600  
Qy 1371 ctgcctctgcctgatggaagccggatgatggcagtggtcgcctcgtgggaaactcttc 1430  
Db 601 CTGCTTCTGCTGATGGAAGCCGGATGATGSCAGTGGCTGCGCTCGTGGTGAACCTTC 660  
Qy 1431 acagccagtgagaggtcgtccctctcctctctctctctctctctctctctctctctctca 1490  
Db 661 ACAGCCAGTGGCAGTGTGCT 720  
Qy 1491 tgagtttgctcacgtgatgcacatcagattgtgcacagactgatttgcacagatttagcgg 1550  
Db 721 TGAGTTTGGTCACGTGATGCATCAGATTTGTGCACAGACTGATTTTGCAGATTTAGCGG 780  
Qy 1551 acaaatgtgaaactgacttcttagaggtgcccacgcgaatgcttgaaaattgggtgtg 1610  
Db 781 AACAAATGTGAAACTTGACTTTGTAGAGTGCCATCGCAAAATGCTTGAATAATTGGGTGTG 840  
Qy 1611 ggaagtcgattccctccgaagattgtcaaaacattataaagatggaagccctcattgcaga 1670  
Db 841 CGAGCTCATCTCCCTCCGCAAGATGTCAAAACATTTATAAAGATGGAAGCCCTATTGCGAGA 900  
Qy 1671 cgatcgtgttgaaaaaacttggcttcttagcttggtgcacacaggtctcttcgacccctgcg 1730  
Db 901 CGATCTGCTTCAAAAACCTTGTGCTTCTAGCTGTGTCAACACAGGTCTTCTGACCCCTGCG 960  
Qy 1731 ccagattgttttgaggaagattgatcagctctctctcaaccacacacatcgctggatgcgc 1790  
Db 961 CCAGATTCTTTTGACAAAGTTGATCAGTCTCTCATACCAACACATCGCTGGATGCTGCTG 1020  
Qy 1791 aagtgaatatccaaatactactcagaaataattaggagttgcagctactccagggcacaaa 1850  
Db 1021 AAGTGAATATGCCAAATACTGCTCAGAAATATTAGGAGTTGCAGCTACTCCAGGCACAAA 1080  
Qy 1851 tatccagctacaccttggacatttggcagggggatacagatggccaataattatggatatct 1910  
Db 1081 TATGCCAGCTACCTTTGGACATTTGGCAGGGGATACCATGGCCAAATATTATGGATATCT 1140  
Qy 1911 ttgagtgaaagtatttccatcgatatatttcaagctgtttttaaagaaagaggataat 1970





```
Db 907 TCGCTGGATGCTGCAAGTGAATATGCCAAATACTCTCAGAAATATTAGGAGTTGCAGCT 966
QY 1837 actccaggcacaataatgccagctacctttggacatttggcaggggatacagatggccaa 1896
Db 967 ACTCCAGGCACAAATATGCCAGCTACCTTTGGACATTTGGCAGGGGGATACGATGCCCAA 1026
QY 1897 tattatgatatctttggagtgaagtattttccatggatatgttttacagctgtttttaa 1956
Db 1027 TATTATGGATATCTTTGGAGTGAAGTATTTCCATGGATATGTTTTACAGCTGTTTTAAA 1086
QY 1957 aaagaaggataataatgaatccagaggttggaaatgaatacagaaacctaatcctgaaacct 2016
Db 1087 AAAGAAGGGATAATGAATCCAGAGTTGGAATGAAATACAGAAACCTAATCCTGAAACCT 1146
QY 2017 ggggatctctggacgcatggacatgctccacaatttcttgaacgtgagccaaaccaa 2076
Db 1147 GGGGATCTCTGGACGGCATGGACATGCTCCACAATTTCTTGAACGTGAGCCAAACCAA 1206
QY 2077 aaagcgttcctaatagtagagggcctgcatgctycgtga 2115
Db 1207 AAAGGTTCTTAATCAGTAGAGGCCCTGCATGCTCGGTGA 1245
```

Search completed: January 14, 2002, 20:25:35  
Job time: 5144 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:56:01 ; Search time 106.63 Seconds  
(without alignments) 965.729 Million cell updates

Title: US-09-833-782-2  
 Perfect score: 3668  
 Sequence: 1 MIARCLLAVRSUURVGGSR.....FLKEEPNOVAFI MSGLHAP 704

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 473505 seqs, 14627329 residues  
Total number of hits satisfying chosen parameters: 473505

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :
SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
```

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	3668	100.0	704	4	Q9BYT8	homo sapien
2	3441	93.8	745	6	P79433	sus scrofa
3	2640	72.0	607	4	Q9BQD0	homo sapien
4	2355	64.2	447	4	Q9ULU4	homo sapien
5	2299	62.7	687	11	Q9EPX1	mus musculus
6	2291.5	62.5	685	13	Q9PTV2	Q9pex1 mus musculus
7	1703	46.4	480	4	Q9BW75	homo sapien
8	1177	32.1	716	3	P97996	Q9bw75 xenopus lae
9	1017.5	27.7	998	10	Q9FXD5	Q9pex1 mus musculus
10	848	23.1	680	2	Q9KVF8	Q9bw75 xenopus lae
11	804	21.9	680	2	Q9LA87	Q9bw75 xenopus lae
12	794.5	21.7	680	2	Q9I765	Q9bw75 xenopus lae
13	781.5	21.3	679	2	Q9CWX1	Q9bw75 xenopus lae
14	702	19.1	678	2	Q9K1E2	Q9bw75 xenopus lae
15	694	18.9	678	2	Q9JX57	Q9bw75 xenopus lae
16	690.5	18.8	713	2	P74571	Q9bw75 xenopus lae
17	681.5	18.6	714	10	Q9JLSL3	Q9bw75 xenopus lae
18	671.5	18.3	685	2	Q9RTU7	Q9bw75 xenopus lae
19	635.5	17.3	723	10	Q9IXA6	Q9bw75 xenopus lae

## ALIGNMENTS

[illegible]

QY 241 ITLKYPHYFPMKKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKVKAKLGYSTHA 300  
DB 241 ITLKYPHYFPMKKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKVKAKLGYSTHA 300  
QY 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGEAREFIFLNKKKECKDRGEYDGKINAWD 360  
DB 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGEAREFIFLNKKKECKDRGEYDGKINAWD 360  
QY 361 LYVYMTQTEELKYSIDQEFLEKPEFVETVTEGLNTYQELLGLSFEQMTDAHVNKSVTL 420  
DB 361 LYVYMTQTEELKYSIDQEFLEKPEFVETVTEGLNTYQELLGLSFEQMTDAHVNKSVTL 420  
QY 421 YTVKDKATGEVLGYLDLXPREGKYHNAACFGLQPGCLLPDGRMMAVAALVNFVSPV 480  
DB 421 YTVKDKATGEVLGYLDLXPREGKYHNAACFGLQPGCLLPDGRMMAVAALVNFVSPV 480  
QY 481 AGPSLLRHDEVRTYFHEFGVHMQICAOQDFARFSGTNTVETDFVEVPSQMLENNWVDVD 540  
DB 481 AGPSLLRHDEVRTYFHEFGVHMQICAOQDFARFSGTNTVETDFVEVPSQMLENNWVDVD 540  
QY 541 SLRSLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEY 600  
DB 541 SLRSLSKHYKDGSPADLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEY 600  
QY 601 AKYCEILGVAATPGTNNPATFGLAGYDQYGYGLWSEVFSFDMFYSCFKKEGINNPE 660  
DB 601 AKYCEILGVAATPGTNNPATFGLAGYDQYGYGLWSEVFSFDMFYSCFKKEGINNPE 660  
QY 661 VGMKYRNILKPGSLDGMMLHNFLEKREPNOKAFMLSRGLHAP 704  
DB 661 VGMKYRNILKPGSLDGMMLHNFLEKREPNOKAFMLSRGLHAP 704

RESULT 2  
P79433 PRELIMINARY; PRT; 745 AA.  
AC P79433;  
DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE ENDOPEPTIDASE 24.16 (EC 3.4.-.-).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
ON NCBI\_TaxID=9823;  
RX [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=LIVER;  
RC MEDLINE=97326108; PubMed=9182559;  
RA Kato A., Sugitara N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;  
RT "targeting of endopeptidase 24.16 to different subcellular  
RT compartments by alternative promoter usage."  
RL J. Biol. Chem. 272:15313-15322(1997).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
DR EMBL; AB000172; BAA19064.1; -  
DR EMBL; AB000173; BAA19065.1; -  
DR EMBL; AB000174; BAA19066.1; -  
DR EMBL; AB000175; BAA19067.1; -  
DR EMBL; AB000425; BAA19105.1; -  
DR EMBL; AB000411; BAA19105.1; JOINED.  
DR EMBL; AB000414; BAA19105.1; JOINED.  
DR EMBL; AB000415; BAA19105.1; JOINED.  
DR EMBL; AB000416; BAA19105.1; JOINED.  
DR EMBL; AB000417; BAA19105.1; JOINED.  
DR EMBL; AB000418; BAA19105.1; JOINED.  
DR EMBL; AB000419; BAA19105.1; JOINED.  
DR EMBL; AB000420; BAA19105.1; JOINED.  
DR EMBL; AB000421; BAA19105.1; JOINED.  
DR EMBL; AB000422; BAA19105.1; JOINED.  
DR EMBL; AB000423; BAA19105.1; JOINED.  
DR EMBL; AB000424; BAA19105.1; JOINED.

EMBL; AB000170; BAA19061.1; -  
EMBL; AB000171; BAA19062.1; -  
EMBL; AB000172; BAA19063.1; -  
EMBL; AB000425; BAA19106.1; -  
EMBL; AB000414; BAA19106.1; JOINED.  
EMBL; AB000415; BAA19106.1; JOINED.  
EMBL; AB000416; BAA19106.1; JOINED.  
EMBL; AB000417; BAA19106.1; JOINED.  
EMBL; AB000418; BAA19106.1; JOINED.  
EMBL; AB000419; BAA19106.1; JOINED.  
EMBL; AB000420; BAA19106.1; JOINED.  
EMBL; AB000421; BAA19106.1; JOINED.  
EMBL; AB000422; BAA19106.1; JOINED.  
EMBL; AB000423; BAA19106.1; JOINED.  
EMBL; AB000424; BAA19106.1; JOINED.  
InterPro: IPR001130; Zn\_MTPeptidse.  
InterPro: IPR001567; Peptidase\_M3.  
Pfam: PF01432; Peptidase\_M3; I.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Metalloprotease; Hydrolase; Zinc.  
SQ SEQUENCE 745 AA; 85109 MW; EC451B8EAD4AC28F CRC64;

Query Match 93.88; Score 3441; DB 6; Length 745;  
Best Local Similarity 95.18; Pred. No. 1.3e-214;  
Matches 656; Conservative 15; Mismatches 19; Indels 0; Gaps 0;

\* QY 15 VGGSRILLRMTGLREVMSPIQAMSSYTVAGRNVLRWDLSPQIKTRTEELIVQTKQYDA 74  
DB 56 VGGSGILLRMTGLREMSPIQAMSSYTVAGRNVLRWDLSPQIKTRTEELIVQTKQYDD 115  
QY 75 VMLGIEBVTYENCLOALADVEVKYIVERTMLDPQHVSSDKVEVRAAASADKRLSRPDI 134  
DB 116 IGLMDIEBVTYENCLOALADVEVKYIVERTMLDPQHVSSDKVEVRAAASADKRLSRPDI 175  
QY 135 EMSMRGDIPIRIVHQLQETCDLGKIKPEARRYLEKSIKMGKENGHLHPQVQNEIKSMKKR 194  
DB 176 EMSMRDIFLRIVRLKETCDLGKIKPEARRYLEKSVKMGKRNGLHLPQVQNEIKAMKKR 235  
QY 195 MSELCIDFNKLNEDDTFLVFSKALGALPDFIDSLSEKTDDEKDKYKITLKYPHYFPVMMK 254  
DB 236 MSELCIDFNKLNEDDTFLVFSKALGALPDFIDSLSEKTDDEKDKYKITLKYPHYFPVMMK 295  
QY 255 CCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKVKAKLGYSTHADFLVLEMTAKSTSR 314  
DB 296 CCIPETRRRMEMAFNTRCKEENTIILOQLPLRAKVAKLLGYSTHADFLVLEMTAKSTHH 355  
QY 315 VTAFDLDLSOKLPLGEAREFIFLNKKKECKDRGEYDGKINAWDLVYVYMTQTEELKYS 374  
DB 356 VTAFDLDLSOKLPLGEAREFIFLNKKKECKDRGEYDGKINAWDLVYVYMTQTEELKYS 415  
QY 375 IDQEFLEKPEFVETVTEGLNTYQELLGLSFEQMTDAHVNKSVTLTYVKDKATGEVLGQ 434  
DB 416 VQDQILKEFPIEVVTEGLNTYQELLGLSFEQMTDAHVNKSVTLTYVKDKATGEVLGQ 475  
QY 435 FYLDLYPREGKYHNAACFGLQPGCLLPDGRMMAVAALVNFVSPVAGRPRLRHDEVRT 494  
DB 476 FYLDLYPREGKYHNAACFGLQPGCLLPDGRMMSVAALVNFVSPVAGRPRLRHDEVRT 535  
QY 495 YFHEFGVHMQICAOQDFARFSGTNTVETDFVEVPSQMLENNWVDVSLRSLSKHYKDGSP 554  
DB 536 YFHEFGVHMQICAOQDFARFSGTNTVETDFVEVPSQMLENNWVDVSLRSLSKHYKDGSP 595  
QY 555 IADDLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEYAKYCEILGVAATP 614  
DB 596 ITDDLLEKLVASRLVNTGLTLRQIVLSKYVDQSILHTNTSLDAASEYAKYCEILGVAATP 655  
QY 615 GTNNPATFGLAGYDQYGYGLWSEVFSFDMFYSCFKKEGINNPEVGMKYRNILKPGG 674  
DB 656 GTNNPATFGLAGYDQYGYGLWSEVFSFDMFYSCFKKEGINNPEVGMKYRNILKPGG 715  
QY 675 SLDGMMLHNFLEKREPNOKAFMLSRGLHAP 704  
DB 715 SLDGMMLHNFLEKREPNOKAFMLSRGLHAP 704

Ds 716 SLDGMDLQNLKREPNOAKFLMSRGLHAP 745

```
RESULT 3
Q9BQD0 PRELIMINARY; PRT; 607 AA.
AC Q9BQD0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:3978) (SIMILAR TO THIMET OLIGOPEPTIDASE 1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004985; AAH04985.1;
DR EMBL; BC001644; AAH01644.1;
SQ SEQUENCE 607 AA; 70304 MW; 1CB7DEAF29A6B2C3 CRC64;

Query Match 72.0%; Score 2640; DB 4; Length 607;
Best Local Similarity 94.8%; Pred. No. 7.1e-163;
Matches 514; Conservative 4; Mismatches 14; Indels 10; Gaps 2;

QY 1 MTARCLAVRSRRVGGSKILLRMTLGRVMSPLQAMSYTVAGRNVLKWLDSPEQIKTR 60
Db 1 MIARCLAVRSRRVGGSKILLRMTLGRVMSPLQAMSYTVAGRNVLKWLDSPEQIKTR 60

QY 61 TEELIVQTKOVYDVGMLGIEVTVENCLOALADYEVKVIYVERTMLDPQHVSSDKEYRA 120
Db 61 TEELIVQTKOVYDVGMLGIEVTVENCLOALADYEVKVIYVERTMLDPQHVSSDKEYRA 120

QY 121 ASTEADKRLSRPDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYLEKSIMKGRNGLHL 180
Db 121 ASTEADKRLSRPDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYLEKSIMKGRNGLHL 180

QY 181 PQVONEIKSMKRMSELCDIDPNKLNEDDTPLVFSKAEGLALPDFIDSLKTDKDYK 240
Db 181 PQVONEIKSMKRMSELCDIDPNKLNEDDTPLVFSKAEGLALPDFIDSLKTDKDYK 240

QY 241 ITLKYPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHA 300
Db 241 ITLKYPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHA 300

QY 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGEAEREFILNKKKECKDRGFYDVGKINAWD 360
Db 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGEAEREFILNKKKECKDRGFYDVGKINAWD 360

QY 361 LYVMTQTEELKYSIDQBFLEKFFPIEVVTEGLLNTYQELLGLSFQMTDAHVWNKSVTL 420
Db 361 LYVMTQTEELKYSIDQBFLEKFFPIEVVTEGLLNTYQELLGLSFQMTDAHVWNKSVTL 420

QY 421 YTVKDKATGEVLGOYLDLYPREGYNHAACGLPGCLLPDGRMMVAALVNFPSQPV 480
Db 421 YTVKDKATGEVLGOYLDLYPREGYNHAACGLPGCLLPDGRMMVAALVNFPSQPV 480

QY 481 AGRPSLLRHDEVRTYFHFGVHMHQICAQTFARFSGTNVETDF-----VEVPSQM 531
Db 481 AGRPSLLRHDEVRTYFHFGVHMHQICAQVSEF-PPVNLPISTFFRKLDDCCQVSNRF 539

QY 532 LE 533
Db 540 FQ 541
```

```
RESULT 4
Q9ULJ4 PRELIMINARY; PRT; 447 AA.
AC Q9ULJ4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIAA1226 PROTEIN (FRAGMENT).
GN KIAA1226.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033052; BAA86540.1;
DR InterPro; IPR000130; Z0_MTPeptidse.
DR InterPro; IPR001567; Peptidase_M3.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 447 AA; 50976 MW; 5A1E6DD76C00E040 CRC64;

Query Match 64.2%; Score 2355; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.3e-144;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 PSTRRMEAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHADFLVLENTAKTSRRTA 317
Db 1 PSTRRMEAFNTRCKEENTIILOQLPLRTKVAKLLGYSTHADFLVLENTAKTSRRTA 60

QY 318 FLDDLSQLKPLGEAEREFILNKKKECKDRGFYDVGKINAWDLYYMTQTEELKYSIDQ 377
Db 61 FLDDLSQLKPLGEAEREFILNKKKECKDRGFYDVGKINAWDLYYMTQTEELKYSIDQ 120

QY 378 EPLKFFPIEVVTEGLLNTYQELLGLSFQMTDAHVWNKSVTLTYVKDKATGEVLGOY 437
Db 121 EPLKFFPIEVVTEGLLNTYQELLGLSFQMTDAHVWNKSVTLTYVKDKATGEVLGOY 180

QY 438 DLYPREGYNHAACGLPGCLLPDGRMMVAALVNFPSQPVAGRPSELRLRHDEVRTYFH 497
Db 181 DLYPREGYNHAACGLPGCLLPDGRMMVAALVNFPSQPVAGRPSELRLRHDEVRTYFH 240

QY 498 EFGVHMHQICAQTFARFSGTNVETDFVEVPQMLNWWVDVSLRRLSKHKDGSPIAD 557
Db 241 EFGVHMHQICAQTFARFSGTNVETDFVEVPQMLNWWVDVSLRRLSKHKDGSPIAD 300

QY 558 DLLEKLVASRLVNTGLLRLQIVLSKVQDQSLTNTSLDAASEYAKYCSILGVAATPGTN 617
Db 301 DLLEKLVASRLVNTGLLRLQIVLSKVQDQSLTNTSLDAASEYAKYCSILGVAATPGTN 360

QY 618 MPATGHLAGGYDGOYGYLVSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGGSLD 677
Db 361 MPATGHLAGGYDGOYGYLVSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGGSLD 420

QY 678 GMDLHNLKREPNOAKFLMSRGLHAP 704
Db 421 GMDLHNLKREPNOAKFLMSRGLHAP 447

RESULT 5
Q9EPX1 PRELIMINARY; PRT; 687 AA.
ID Q9EPX1
```

```
AC Q9EPX1; 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15).
GN THOPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20538378; PubMed=10969067;
RA Tullai J.W., Cummins P.M., Pabon A., Roberts J.L., Lopingco M.C.,
RA Shrimpton C.N., Smith A.I., Martignetti J.A., Ferro E.S.,
RA Gluckman M.J.;
RT "The Neuropetide Processing Enzyme EC 3.4.24.15 Is Modulated by
RT Protein Kinase A Phosphorylation.";
RL J. Biol. Chem. 275:36514-36522(2000).
DR EMBL; AF314187; AAC35061.1; -;
DR MGD; MGI:1354165; Thopi.
DR InterPro: IPR001567; Peptidase_M3.
DR InterPro: IPR000130; Zn_MTPeptidse.
DR Pfam: Pf01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 687 AA; 77994 MW; F34A63D7C2A6D018 CRC64;

Query Match 62.7%; Score 2299; DB 11; Length 687;
Best Local Similarity 64.8%; Pred. No. 1e-140;
Matches 425; Conservative 92; Mismatches 139; Indels 0; Gaps 0;

Qy 46 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVVKIVERTM 105
Db 22 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVVKIVERTM 81

Qy 106 LDFQHVSSDEKRAAETADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 165
Db 82 LDFQHVSSPCADIRAAETADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 141

Qy 166 LEKSIKMGKRNGLHLPQVQVNEIKSMKRMSELICIDFNKNLNEDDTFLVFSKAEGLALPD 225
Db 142 LERLIKLRNGLHLPQDTQEKIRKNIKSLICIDFNKNLNEDDTFLVFSKAEGLALPD 201

Qy 226 DFIDSLEKTDGKYNLKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285
Db 202 DFLISLEKADGKILKVTILKYTPHYFPVLLKCHVPETRRLLLEAFNCRCKEENCAILKELVS 261

Qy 286 LRTKVALILGYSTHADFLVLENTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 345
Db 262 LRAQKSSILGFPHTHADFLVLENTAKTSQVATFDELDAQLKPLGEQERAVILEKKEAC 321

Qy 346 KDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKYEFPVEVTEGLLTYQELLGLSF 405
Db 322 AKRGLPFDGRHAWDMRYMNQVEETRYRVDQNLKKEYPFQVYTRGLLTYQELLGLTF 381

Qy 406 EQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLOPGCLLPDQGS 465
Db 382 TLEGAANWHEDEVRLYSVRDAASGEEIGKFFLDLYPREGKYNHAAACFGLOPGCLLPDQGS 441

Qy 466 MMVAALVYNFQSVAGRPVLRHDEVRTYFHERGHVNHQICAOITDFARFSGTINVTDFV 525
Db 442 OLAVAAMVYANFTKTPDAPSLQLHDEVTYFHERGHVNHQICSAERAMFSGTHVERDFV 501

Qy 526 EVPSOMLENWVDVDSLRRLSKHYKDGSP IADDLLEKLVASRLVNTGLLTIRQIVLSKVD 585
Db 502 EAPSOMLNWWYERKEPLRMSSQHYRTGGEAPDQDLLEKLVASRLVNTGLLTIRQIVLSKVD 561

Qy 586 QSLHTNTSLDAASAYKVCSEILGVAATPGTNMPATFGHLAGGYDGYGYGLWSEVFSMD 645
Db 562 QVLTHTQTDADPAEYARLCQELIGVPATPGTNMPATFGHLAGGYDGYGYGLWSEVFSMD 621
```

```
Qy 646 MFYSCKFKEGIMNPVGMYKYNLILKPGSLDGMDFLHNFLEKRPNQKAFILMSRGL 701
Db 622 MFHTFRKQEGVLSKPVGMVYRTSILRPGSSDASAMILKQFLGRDPKQDAFLLSKGL 677

RESULT 6
Q9PTV2 PRELIMINARY; PRT; 685 AA.
AC Q9PTV2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE THIMET OLIGOPEPTIDASE.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC Okida N., Ishikawa K., Tokumoto M., Tokumoto T.;
RT "Xenopus thimet oligopeptidase.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB030904; BAA89309.1; -;
DR InterPro: IPR001567; Peptidase_M3.
DR InterPro: IPR001567; Peptidase_M3.
DR Pfam: Pf01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 685 AA; 78137 MW; F41C5EE45BBB78ED CRC64;

Query Match 62.5%; Score 2291.5; DB 13; Length 685;
Best Local Similarity 64.4%; Pred. No. 3.1e-140;
Matches 426; Conservative 98; Mismatches 135; Indels 3; Gaps 1;

Qy 46 NVLRWDLSPQIKTRTEELIVQTKQVDAVGMGLGIEEVTYENCLOALADVVKIVERTM 105
Db 19 NVLRWDLTPKQIEITLGEVTEQTKRVYDQVGSLLSSVTYDNTLKLVDADVVEVSGKRS 78

Qy 106 LDFQHVSSDEKRAAETADKRLSRDIEMSRGDIPIERIVHLEQTCDLGKIKPEARRY 165
Db 79 LDFQHVSPSEIRAVSTAADKRLSEDFVEQSMREDYVRRIEHLQTIIPSTNLKPEAKRW 138

Qy 166 LEKSIKMGKRNGLHLPQVQVNEIKSMKRMSELICIDFNKNLNEDDTFLVFSKAEGLALPD 225
Db 139 LERVIKLSQRNGLNLPSTQEKIKSIKKNSTLSDFNKNLNEDITFLFTTKEELGLPD 198

Qy 226 DFIDSLEKTDGKYNLKYTPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285
Db 199 DFLNSLEKAGDGKILKVTILKYTPHYFPVLMKKCFIPNTRKROKVEAAFNCRKEENSRILQELVQ 258

Qy 286 LRTKVALILGYSTHADFLVLENTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 345
Db 259 LREGNSITLGFTHADFLVLENTAKTSRVTAFDLDLSQKLPLGEAREFERILNKKKEC 318

Qy 346 KDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKYEFPVEVTEGLLTYQELLGLSF 405
Db 319 KKRGLNFDQINAWDMRYMNQVEETSYNVQDNFLKEFFIEVVTGTLGIYQELLGLNF 378

Qy 406 EQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAAACFGLOPGCLLPDQGS 465
Db 379 VLEGAANWHEDEVRLYSVRDATQGLCKFYLDLYPREGKYNHAAACFGLOPGCLLPDQGS 438

Qy 466 MMVAALVYNFQSVAGRPVLRHDEVRTYFHERGHVNHQICAOITDFARFSGTINVTDFV 525
Db 439 QISVAAMVYANFTKTPDAPSLQLHDEVTYFHERGHVNHQICAOAEVFLFSGTGVERDFV 498

Qy 526 EVPSOMLENWVDVDSLRRLSKHYKDGSP IADDLLEKLVASRLVNTGLLTIRQIVLSKVD 585
Db 499 EAPSOMLNWWYERKEPLRMSSQHYKSGQAIPKEVLEHLIKSRANTGLLNLRQIVLAKID 558

Qy 586 QSLHTNTSLDAASAYKVCSEILGVAATPGTNMPATFGHLAGGYDGYGYGLWSEVFSMD 645
```

Db 559 QVLTQIGVDPVEEYKSLNSNETLIGIPATPGTNMLSAFAHLAGGYDAQYGYLWSEVYSND 618  
Qy 646 MYSCFKKEGINPVGMYKYNRLILKPGSLDGMDLNHLNFKRPNQKAFILMSRGL---H 702  
Db 619 YITRFKQEGINSKKVGADYRNCILKPGSLDGMDLNHLNFKRPNQKAFILMSRGL 678  
Qy 703 AP 704  
Db 679 AP 680  
RESULT 7  
ID Q9BW75 PRELIMINARY; PRT; 480 AA.  
AC Q9BW75;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DE TIMET OLIGOPEPTIDASE 1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=NEUROBLASTOMA;  
RA Strausberg R.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC000583; AA000583.1; -.  
SQ SEQUENCE 480 AA; 54735 MW; 96250597AC877913 CRC64;

Query Match 46.4%; Score 1703; DB 4; Length 480;  
Best Local Similarity 65.5%; Pred. No. 2.4e-102;  
Matches 306; Conservative 68; Mismatches 93; Indels 0; Gaps 0;  
Qy 235 DDYKYLTKLPHYPPVMMKCCIPETRRMEMAFNTRCKEENIILQQLPLRTKVKLL 294  
Db 2 EDGKLKYLKPHYPLKCHVPTRRKVEAFNCRCKEENALIKELVTLRAQKSRLL 61  
Qy 295 GYSTHADFVLENTAKTSRVTAFDDLSQKLPLGEAREPILNKKKECKDRGFYD 354  
Db 62 GPHTHADVLENNMAKTSOTVATFDELQAKLPLGEQERAVILELKRACERRGLP 121  
Qy 355 KINANDLYYMTQTEELKYSIDQELKYEFPVTEGLLNTYQELLGLSPQMTDAHV 414  
Db 122 RIRAWDMRYNNQVEETRYCDONLLKEYFPVQVYTHGLLGYELLGLAFHHEGAS 181  
Qy 415 NKSVTLYTKDKATGEVLGQFVLDLYPREGKYNHAACFGLQPGCLLPDGRMMVAALV 474  
Db 182 HDVRLYTARDAASGEVKGFLVDLYPREGKYGHACFGLQPGCLLRQDSRQIA 241  
Qy 475 NFSQVAGRPSLLRHDEVRTYFHEFGHYMHQICATDFARFSGTNVETDFVEVPSQ 534  
Db 242 NFKPTADAPSLQLHDEVTYFHEFGHYMHQICSAEAFMFGSTHVRDFVEAPSO 301  
Qy 535 WYWDVDSLRRLSKHYKDGSPITADLLEKLVASRLVNTGLTLRLQIVLSKVDS 594  
Db 302 WYWEQEPRLRSRHYRTGSAPVRELEKLIERSQANTGLNLRQIVLAKVQDQAL 361  
Qy 595 DAASFYAKYCSILGVAATPGTNMPATFGHLAGGYDGYGYGLWSEVSDMFYSCPK 654  
Db 362 DPAEYARLCQELVLPATPGTNMPATFGHLAGGYDGYGYGLWSEVSDMFHFR 421  
Qy 655 GIMNPEVGMKYNRLILKPGSLDGMDLNHLNFKRPNQKAFILMSRGL 701  
Db 422 GVLNSKVGMDYRSCILRPGGSDASAMLRRLRFLGRDPKQDAFLLSKGL 468

RESULT 8  
ID P97996 PRELIMINARY; PRT; 716 AA.  
ID P97996

P97996;  
AC 01-JUL-1997 (Tremblrel. 04, Created)  
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE INTRACELLULAR METALLOPROTEINASE MEPP.  
GN MEPP.  
OS Aspergillus fumigatus (Sartorya fumigata).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=5085;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97388579; PubMed=9245813;  
RA Ibrahim-Granet O., d'Enfert C.;  
RT "The Aspergillus fumigatus mepp gene encodes an 82 kDa intracellular metalloproteinase structurally related to mammalian thimet oligopeptidases";  
RL Microbiology 143:0-0(0).  
DR EMBL; U85769; AAB66656.1; -.  
DR MEROPS; M03.009; -.  
DR InterPro; IPR001567; Peptidase\_M3.  
DR InterPro; IPR000130; Zn\_Mrpeptidse.  
DR Pfam; PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
SQ SEQUENCE 716 AA; 81923 MW; 6480A86C90185322 CRC64;

Query Match 32.1%; Score 1177; DB 3; Length 716;  
Best Local Similarity 37.7%; Pred. No. 4.9e-68;  
Matches 273; Conservative 125; Mismatches 253; Indels 74; Gaps 15;  
Qy 33 PLQAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVOTKVYD-AVGMLGTEETVENCLOA 91  
Db 9 PPQAPPLTATAQSI-----DDAKRLIETSRKVRDMIVANKPESATFDSVLAP 58  
Qy 92 LADVEVKYIVERTMLDFQHVSDKEVRAASTADKRLSRFDIEMSRGDI-----ERIVH 148  
Db 59 LAHDENTMALESHILSFYQAVSTEQELRDASSKABELMDEFFIETVMREDVFKLDAVLK 118  
Qy 149 LOETCDLGIKPEARYLEKSIKMKRNLGLHPE-QVQNEIKSMKMRMSELCDIFNKNLN 207  
Db 119 KNESLD-----PESRLLKEKHQYIRNGLGPAGPKDRFKRKEIKRLSQISIEFOKNL 173  
Qy 208 EDTTLVFSKAEALGALPDDFIDSLKTD---DDKYKITLKYPHYPPVMMKCCIPETRRM 264  
Db 174 ENGGINFTRESDGVPDVLGLKKEGEGEKWLTFKYPDLPFTMYAKNAETRRKL 233  
Qy 265 ENAFNTRKEENTIILOQLPLRTKVKLLGYSTHADFVLENTAKTSRVTAFDLDLSQ 324  
Db 234 MIENENKC-NQNVPLFREAVILRDEAARLLGYPNHAAFRIDKMAKTPKTVDDFLGLRS 292  
Qy 325 KKLPLGEAREPILNKKKECKDRGFYDQKINANDLYYMTQTEELKYSIDQELKYE 384  
Db 293 RLTAGGHKEIKALLEKRADESERGEPTDGRYLYMDHREYDRLMLEKDYSLDQQLIAEYF 352  
Qy 385 PLVETGLNTYQELLGLSPQMTDAH-----VNNKSVTLTYV-KDKATGE 430  
Db 353 PLQTTIEGLMKLFEELFGLVFEITGEDREKVAPTKGSDIVWHEDVQVFSVWNDGEGS 412  
Qy 431 -VLGQFYLDLYPREGKYNHAACFGLQPGCLLPDGRMMVAALVNFSPQVAGRPSLLRH 489  
Db 413 GFVGYLILDLFFPRGKYGHAAFNALQPGIDAGKRRYPATALVCNFTKPTAKKPSLLAH 472  
Qy 490 DVRIYFHEFGHYMHQICATDFARFSGTNVETDFVEVPSQMLENNWVDSLRRLSKHY 549  
Db 473 DEWTLFHELGHGCHDLVSRITTSYRFGHTSTVRDFVEAPSQMLENNWCVTPSOLSKSHY 532  
Qy 550 KDGSP-----IADDLLEKLVASRLVNTGLTLRLQIVLSKVDSOHLT 590  
Db 533 STLSPEYLAGWQEQANGKPPQPPERIPDEIENLIRTKHVNDALFNLRQLHFGIFDMTH 592  
Qy 591 NTS-----LDAASEYAKYCS-----EILGVAATPGTNMPATFGHLAGGYDGY 636

Db 593 AESHEQIQKLPSTTYNQLRQKIALQDGPVGLMGDEWG-HGEATFGHLGGYDAGYGY 651  
QY 637 LWSEVFMDFYSCFKKRGINMNVGMYKRNILKPGSLDGMMLHNFLLKRPNQKFL 696  
Db 652 LSSQVSTDMETFTYKDDP-MNPAAGRRYRGVLEKSGSQDEMKTLDFTLGRKPTDAFY 710  
QY 697 MSRGL 701  
Db 711 KEGL 715

RESULT 9  
Q9FXD5 PRELIMINARY; PRT; 998 AA.  
AC Q9FXD5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE F12A21.16.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC F12A21 from chromosome  
I.,"  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,  
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,  
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
RA Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N., Theologis A.,  
RA Ecker J.;  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Ecker J.R.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,  
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,  
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
RA Theologis A., Ecker J.;  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP EMBL; AC08113; AAC28905.1; -  
DR InterPro: IPR001567; Peptidase\_M3.  
DR InterPro: IPR000130; Zn\_MTPeptidse.  
DR Pfam: PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.

SQ SEQUENCE 998 AA; 112949 MW; 49CBE1C8225955E8 CRC64;  
Query Match 27.7%; Score 1017.5; DB 10; Length 998;  
Best Local Similarity 33.7%; Pred. No. 1.7e-57;  
Matches 221; Conservative 157; Mismatches 244; Indels 33; Gaps 9;  
QY 51 DLSPEQIKTRTEELIVQTKQVYDAGVLMGIBEVTYENCQAQADAVEVKYIIVERTMLDPQ 110  
Db 40 NLSAKEILDAAEIIHKSTRVHDVALVSLDKSYENVVPLAELEARQLSLIOCCVFPK 99  
QY 111 HVSSDKVEVRAASTEADKRLSRFDIEMSRMGDIFERIVHLQETCDLGIKIKPARYLEKSI 170  
Db 100 MSLPHDNVRKASTAEQKIDAHILSCRREDVY-RIIKIY-AARGESISPAAKYLQCLV 157  
QY 171 KMGKRGHLHPEQVQNETSKMKRWSLCLIDFNKNLNDDEDFELVFSKALGALPDDFTDS 230  
Db 158 RFDNGNLNLTAIKREVERUKYIEIDLSRYIQNLNEDSSCLFFTEDELGLPLEFLQN 217  
QY 231 LEKTDGDKYKITLKYPHYFPVMMKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKV 290  
Db 218 LEKTONKEFLTLSESRHVAALTELCKIAKTRKTVMAYGKRCGDTNPVQLRQVSRHRL 277  
QY 291 AKLGYSTHADFLVLEMTAKSTSVTAFLDLSOKLPLGSEAREFRLNLKKCKECKDRGF 350  
Db 278 ACVGYAHFADYALDRRMSKTSMRVIRFLEDISSLTDLAIREFSILEDLKKE----- 331  
QY 351 EYDGKI--NAWDLYYYTOTEELKYSIDOEFLKEYVEVTVTEGLLNTYQELLGLSPEOM 408  
Db 332 --EGEIPFGVEDLLYYIKRVEELQDLDFDIRQYFPVNLVLSGFIKICQDLFGIKFEV 389  
QY 409 TDAHWKNSVTLYTVKDKATGEVLGQFVLDLYPREGYNHAAACFGLQPGCLLPDGSRRMA 468  
Db 390 TEVDVWYHDIRAFVDFDSGSGKLIGYFLDMFTREGKNCVVALQNNALFSGACQIP 449  
QY 469 VAALVNFSPQVAGRPSLLRHDEVRTYFHEFGHYHMOICAGTDFARSGTNVETDFVEVP 528  
Db 450 VALLIAQFAKDGSGEAVPLGFSVYVNLHFHEFGHYHMOICNRSASFSGLRVDPDFEIP 509  
QY 529 SOMLENWMDVDSLRRLSKHYKDGSPADLLLEKLVASRLVNTGLTLRQIVLSKVQDSL 588  
Db 510 SOLLENW-----QDITK-----PLVDEVCKTLKWRYSFSAKLSQELIYCLFDQII 556  
QY 589 HTNTSLD-----AASEYAKYCEILGVAATPGTNMPATFGHLAGGYDGOYGYLWSEVPSM 644  
Db 557 YSDDDADLLQLIRSLHPKV---MIGLPVVECTNPASCFPRAVIGSEATCYSLRWSEVAA 613  
QY 645 DMFYSCKFKKGGIMNPEVGMKYNRLILAPGSLDGMMLHNFLLKRPNQKAFELMSR 699  
Db 614 DIFASKF-GDGHPLNYAGLOFRDKVLAPGGGKPEMLLTNLFGLRPSPTQAFIASR 667

RESULT 10  
Q9KVF8 PRELIMINARY; PRT; 680 AA.  
AC Q9KVF8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE OLIGOPEPTIDASE A.  
GN VC0188.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=566;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,









Query Match		18.9%	Score 694;	DB 2;	Length 678;
Best Local Similarity		28.2%	Pred. No. 8.5e-37;		
Matches 190;		Conservative 131;	Mismatches 309;	Indels 44;	Gaps 18;
Qy	55	EQIKRTEELIVQT--KQVYDAVGMLGIEEVT-YENCLOALADVEVYIIVERT-----ML	106		
Db	16	DOIKTEIDIKPALQTAIAEAEREQIAAIAQAQTHGTWANTVEPLTG-----ITERVGRIGWVV	70		
Qy	107	DFPHVSSDKVEVRAASTADKRLSRFDIEMSMRGDIFERIVHLOETCDLGKIRPEARRYL	166		
Db	71	SHLSNVTDTPELRAAYNELMPEITVFTEIGQDIELYNRFKTIKNSPFDTLSHAQTKL	130		
Qy	167	EKSIKMGKRNGLHPEQVQNEIKSMKKRMSLCIDFNKN-LNEDDTELVF--SKAELGAL	223		
Db	131	NHDLURDFVLSGAELPPQQAELEAKLQTEGQOLSAKFSQNVLDATDAFGIYFDDAPLAGI	190		
Qy	224	PDD----FIDSLEKTDDDKYKITLKYPHYFPVMKKCCIPETRRMEMAFNTRCKE-----	274		
Db	191	PEDALAMFAAAQSEGKTYKIGIQIPHYLAVIQYADNRKLRQIYRAYVTRASELSDDG	250		
Qy	275	--ENTIILQQLLPRTKVKALLGYSTHADVFLEMTAKSTSRVTAFLDDLSQKLPLGEA	332		
Db	251	KFDNTANIDRTLENALQTAKLILGFKNYAELSLATKMDTPEQVLNLFHDLARRAKPYAEK	310		
Qy	333	EREFLNLKKKECKDRGFEDGKINADLYYYMTQTTELKYSIDQEFLEKYEFPTEVVTEG	392		
Db	311	D--LAEVKAFARESLGL---ADLPWDLGYAGEKLEAKYAFSETEVKYFPVKVGLNG	364		
Qy	393	LINTYQELLGLSFEQMTDAHWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAACF	452		
Db	365	LFAQIKKLYIGIGFTEKT-VPVWHKDVRYFELQ--NGETIGGVYMDLYAREGKRGGAWN	421		
Qy	453	GLQPGCLLPDGSRRMAVAALVNVFSQPVAGRPSSLRLRHDEVTYEFHFGHVMHQICAQTD	512		
Db	422	DYKGRRRFSDGTQLQPTAYLCNFTPPVGGKEARLSHDEILTLFHTGCHGLHLLTQVDE	481		
Qy	513	ARFSGTN-VETDFEVPSPOMLENWVDVDSLRRLSKHYKDGSPADDDLEKLVASRLVNT	571		
Db	482	LGVSGINGVEWDVVELFSQFNFVWEYNVLAQMSAHEETGVPLPKELFDKMLAAKNFOR	541		
Qy	572	GLLTIRQIVLSKVQSLHNTSLDAASEYAKYCSFI---LGVAATPGTNMPA-TFGHL-A	626		
Db	542	GMELVRQMEFALFDMMIYSEDEGRLKNWQVLDVSRKEVAVVRPPEYNRRFANSFGHIFA	601		
Qy	627	GGYDQYGYGLWSEVFSMDMYSCFKKEGINPEVGMKYRNLIILKPGSLDGMMLHNF	686		
Db	602	GGYSAGYYSYAWAEVLSADA-YAAFEESDDV-AATGKRFWQEIILAVGGSRSAAESFKA	659		
Qy	687	KREPQKAFMSRG	700		
Db	650	GREPSIDALLRHSG	673		

Search completed: January 14, 2002, 19:02:18  
Job time: 377 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:02:21 ; Search time 198.95 seconds  
(without alignments)  
9114.063 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcgcggcgccctttt.....gagccctgcatgctygtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2109.4	99.7	2864	22 AAF84499	Human aminopeptidase
2	1590.8	75.2	2148	13 AAQ27307	Rabbit endopeptidase
3	1178.6	55.7	1658	22 AAH14379	Human CDNA sequenc
4	790.8	37.4	2070	13 AAQ24875	Human amyloidin pr
5	618.4	29.2	776	22 AAH06280	Human CDNA clone (
6	339	16.0	339	21 AAH31048	Human colon cancer
7	308.8	14.6	900	21 AAC98797	Human pancreatic c
8	117	5.5	1830121	17 AAT42083	Haemophilus influe
9	112.8	5.3	936	22 AAF58252	Oligonucleotide D1
10	112.8	5.3	936	22 AAF58254	Oligonucleotide D1
11	112.8	5.3	936	22 AAF58257	Oligonucleotide D1

12	112.8	5.3	936	22 AAF58259	Oligonucleotide D2
13	112.8	5.3	936	22 AAF58262	Oligonucleotide D2
14	112.8	5.3	938	22 AAF58255	Oligonucleotide D1
15	111.8	5.3	936	22 AAF58252	Oligonucleotide D1
16	111.8	5.3	936	22 AAF58254	Oligonucleotide D1
17	111.8	5.3	936	22 AAF58257	Oligonucleotide D1
18	111.8	5.3	936	22 AAF58259	Oligonucleotide D2
19	111.8	5.3	936	22 AAF58262	Oligonucleotide D2
20	111.8	5.3	938	22 AAF58255	Oligonucleotide D1
21	94.8	4.5	588	22 AAH11342	Human CDNA clone (
22	75	3.5	2037	21 AAF21572	N. meningitidis pa
23	75	3.5	2037	21 AAA81287	N. meningitidis pa
24	75	3.5	2037	21 AAZ53270	Neisseria meningit
25	75	3.5	2037	21 AAZ54604	Neisseria meningit
26	75	3.5	5253	21 AAA81478	N. meningitidis pa
27	75	3.5	349980	21 AAF21544	Neisseria meningit
28	75	3.5	1437668	21 AAA81490	N. meningitidis B
29	73.4	3.5	1035	21 AAZ53267	Neisseria meningit
30	73.4	3.5	1035	21 AAZ54601	Neisseria meningit
31	73.4	3.5	1431	21 AAF21569	N. meningitidis pa
32	73.4	3.5	1431	21 AAA81284	N. meningitidis pa
33	71.2	3.4	623	21 AAF07718	Fusarium venenatum
34	70.2	3.3	2037	21 AAF21571	N. meningitidis pa
35	70.2	3.3	2037	21 AAF21574	N. meningitidis pa
36	70.2	3.3	2037	21 AAA81286	N. meningitidis pa
37	70.2	3.3	2037	21 AAA81289	N. meningitidis pa
38	70.2	3.3	2037	21 AAZ53268	Neisseria meningit
39	70.2	3.3	2037	21 AAZ53271	Neisseria meningit
40	70.2	3.3	2037	21 AAZ54603	Neisseria meningit
41	70.2	3.3	2037	21 AAZ54606	Neisseria meningit
42	67.4	3.2	465	22 AAI14810	Probe #4743 for ge
43	67.4	3.2	465	22 AAI36168	Probe #4854 used t
44	67.4	3.2	465	22 AAI04602	Probe #4593 used t
45	66.4	3.1	2382	21 AAC50567	Arabidopsis thalia

#### ALIGNMENTS

RESULT	1
AAF84499	
ID	AAF84499 standard; cDNA; 2864 BP.
XX	
AC	AAF84499;
XX	
DT	25-JUL-2001 (first entry)
XX	
DE	Human aminopeptidase 22196-encoding cDNA.
XX	
KW	Human: aminopeptidase 22196; neutral zinc metalloproteinase;
KW	metalloenzyme; analgesic; cytostatic; lung cancer; colon cancer; tumour;
KW	pain; drug screening; spleen disorder; infectious disease;
KW	immune disorder; lung disorder; colon disorder; liver disorder;
KW	uterine disorder; brain disorder; skin disorder; cardiac disorder;
KW	vascular disorder; bone disorder; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
CDS	300..2414
FT	/*tag= a
FT	/product= ""Human aminopeptidase 22196"
XX	
PN	WO200123590-A2.
XX	
PD	05-APR-2001.
XX	
PF	02-OCT-2000; 2000WO-US27214.
XX	
PR	30-SEP-1999; 99US-0409180.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	

PI Kapeller-Libermann R, White D, Silos-Santiago I;

XX WPI: 2001-300101/31.

DR P-PSDB: AAB73485.

XX Novel human aminopeptidase, used to treat colon and lung cancer, and to  
PT identify modulators used to treat e.g. cirrhosis -

XX Claim 3; Fig 1A-C; 109pp; English.

CC This sequence represents cDNA encoding a novel human aminopeptidase,  
CC designated 22196. cDNA clones encoding human aminopeptidase 22196  
CC were identified in a bone marrow cDNA library using an EST (expressed  
CC sequence tag) selected on the basis of homology to aminopeptidase  
CC sequences, and were assembled to give the full-length cDNA.  
CC Aminopeptidase 22196 is thought to be a metalloenzyme, as it contains  
CC a zinc-binding region signature found in neutral zinc metalloproteinases.  
CC Expression analysis indicated that the aminopeptidase is highly  
CC expressed in normal osteoblasts, testis, skeletal muscle, foetal kidney  
CC and foetal liver. It is also expressed in normal breast, lung, and colon  
CC tissue, and is overexpressed in lung and colon cancer. Expression was  
CC found to be downregulated during stromal cell osteoblast lineage  
CC maturation. Aminopeptidase 22196 may be used to treat pain, and cancers  
CC of the lung and colon. It can also be used in drug screening to identify  
CC compounds which bind to it and modulate its activity. Such compounds may  
CC be used in the treatment of a wide variety of conditions, including  
CC spleen disorders (e.g., splenomegaly); infectious diseases (e.g.,  
CC malaria, tuberculosis); immune disorders (e.g., rheumatoid arthritis,  
CC systemic lupus erythematosus, acquired immunodeficiency syndrome (AIDS));  
CC lung disorders (e.g., pulmonary embolism, emphysema, bronchial asthma);  
CC colon disorders (e.g., diarrhoea, Crohn's disease, ulcerative colitis);  
CC liver disorders (e.g., hepatic cirrhosis, hepatitis); uterine disorders  
CC (e.g., endometriosis); brain disorders (e.g., meningitis, Alzheimer's  
CC disease, Parkinson's disease, Huntington's disease); skin disorders  
CC (e.g., vitiligo, urticaria, skin tumours, psoriasis); cardiac and  
CC vascular disorders (e.g., myocardial infarction, chronic ischaemic heart  
CC disease, hypertension, atherosclerosis); and bone disorders (e.g.,  
CC osteoporosis, Paget's disease). Nucleotides encoding aminopeptidase 22196  
CC may be used in gene therapy.

XX Sequence 2864 BP; 854 A; 556 C; 693 G; 761 T; 0 other;

Query Match 99.7%; Score 2109.4; DB 22; Length 2864;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2110; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 atgacgcccgtgcttttggctgtgcagagcctccgagagttggtgtccaggatt 60

Db 300 atgacgcccgtgcttttggctgtgcagagcctccgagagttggtgtccaggatt 359

Qy 61 ttactcagaatgacgttaggaagaagatgctctctcttcagcgaatgtcttcctat 120

Db 360 ttactcagaatgacgttaggaagaagatgctctctcttcagcgaatgtcttcctat 419

Qy 121 actggtgctgcagaaatgttttaagaaggatctttccaccagagcaataaaacaaga 180

Db 420 accgtgctgcagaaatgttttaagaaggatctttccaccagagcaataaaacaaga 479

Qy 181 actgaggagctcattgtgcagaccacaaaggtgtacgatgtgttggaatgctcggtatt 240

Db 480 actgaggagctcattgtgcagaccacaaaggtgtacgatgtgttggaatgctcggtatt 539

Qy 241 gaggaaagtaacttcagaaactgtctgcagacactgagcagatgtagaagtaaaagtatata 300

Db 540 gaggaaagtaacttcagaaactgtctgcagacactgagcagatgtagaagtaaaagtatata 599

Qy 301 gtggaaggaccatgctagactttccacagatgtatcctctgcacaaagaagtagcagca 360

Db 600 gtggaaggaccatgctagactttccacagatgtatcctctgcacaaagaagtagcagca 659

Qy 361 gcaagtacagaagcacaagaactttctcgtttttgatttattgagatgagcatgagagga 420

Db 1740 gcaagtacagaagcacaagaactttctcgtttttgatttattgagatgagcatgagttggt 1799

Db 660 gcaagtacagaagcacaagaactttctcgtttttgatttattgagatgagcatgagagga 719

Qy 421 gatatattgagaagaattgttcatttacagaaacacctgtgatctctggggaagataaaacct 480

Db 720 gatatattgagaagaattgttcatttacagaaacacctgtgatctctggggaagataaaacct 779

Qy 481 gaggccgacgactactctggaaaagtcaattaaaatgggggaaaaaagaatgggtccactctt 540

Db 780 gaggccgacgactactctggaaaagtcaattaaaatgggggaaaaaagaatgggtccactctt 839

Qy 541 cctgaacaagtacagaatgaaatcaaatcaatgaagaaaaagaatgagtgactgtatt 600

Db 840 cctgaacaagtacagaatgaaatcaaatcaatgaagaaaaagaatgagtgactgtatt 899

Qy 601 gattttaacaaaaacctcaatgagatgataccctctctgttatttccaaggctgaactt 660

Db 900 gattttaacaaaaacctcaatgagatgataccctctctgttatttccaaggctgaactt 959

Qy 661 ggtctctctctgatgtatttattgacagtttagaaaagacagatgatacaagtataaa 720

Db 960 ggtctctctctgatgtatttattgacagtttagaaaagacagatgatacaagtataaa 1019

Qy 721 attaccttaaatatcacactatttccctgtcatgaagaaatgtgtatccctgaaacct 780

Db 1020 attaccttaaatatcacactatttccctgtcatgaagaaatgtgtatccctgaaacct 1079

Qy 781 aagaagaagtggaaaatggcttttaatacaagggtgcaagagggaacacacataatttg 840

Db 1080 aagaagaagtggaaaatggcttttaatacaagggtgcaagagggaacacacataatttg 1139

Qy 841 cagcagctactcccactgcgcaaccaaggtgccaaaactactcgtttatagcacacatgct 900

Db 1140 cagcagctactcccactgcgcaaccaaggtgccaaaactactcgtttatagcacacatgct 1199

Qy 901 gactctgctctgaaatgaacacatgcacagacacaaagcgcgttaacagcgtttctatag 960

Db 1200 gactctgctctgaaatgaacacatgcacagacacaaagcgcgttaacagcgtttctatag 1259

Qy 961 gatttaagccagaagttaaaaccccttggtggaagcagacagagttattttgaatttg 1020

Db 1260 gatttaagccagaagttaaaaccccttggtggaagcagacagagttattttgaatttg 1319

Qy 1021 aagaaaaaggaaatgcaaaagacaggggttttgaatatgtatgggaaaaatcaatgcctggat 1080

Db 1320 aagaaaaaggaaatgcaaaagacaggggttttgaatatgtatgggaaaaatcaatgcctggat 1379

Qy 1081 ctatattactacatgactcagacagaggaactcaagtattccatagaccaagaggtctctc 1140

Db 1380 ctatattactacatgactcagacagaggaactcaagtattccatagaccaagaggtctctc 1439

Qy 1141 aaggaataacttcccaattgaggtgctcactgaaagccttgctgaacacactaccaggagttg 1200

Db 1440 aaggaataacttcccaattgaggtgctcactgaaagccttgctgaacacactaccaggagttg 1499

Qy 1201 ttgggaacttctatttgaacaaatgacagatgctcatgtttggaaacaaagaggttacacctt 1260

Db 1500 ttgggaacttctatttgaacaaatgacagatgctcatgtttggaaacaaagaggttacacctt 1559

Qy 1261 tatactgtgaaggataaaagctacaggaagaagtattgggacagattctatttggacctctat 1320

Db 1560 tatactgtgaaggataaaagctacaggaagaagtattgggacagattctatttggacctctat 1619

Qy 1321 ccaaggggaagaaaaatacaatcatcgccctgcttcctcggtctccagcctggtgcctcttg 1380

Db 1620 ccaaggggaagaaaaatacaatcatcgccctgcttcctcggtctccagcctggtgcctcttg 1679

Qy 1381 cctgaatgaagcccgagatgattggcagtggtccctcgtgggaaacttctcacagcagatg 1440

Db 1680 cctgaatgaagcccgagatgattggcagtggtccctcgtgggaaacttctcacagcagatg 1739

Qy 1441 gcaggtcgtcctctcctctgagacacagcaggtgagagacttactttcagagtttgggt 1500

Db 1740 gcaggtcgtcctctcctctgagacacagcaggtgagagacttactttcagagtttgggt 1799

```

QY 1501 cagtgatgacatcagatttgcacagactgattttgcagctatttagcggagcaaaatgtg 1560
DB 1800 cagtgatgacatcagatttgcacagactgattttgcagctatttagcggagcaaaatgtg 1859
QY 1561 gaaactgactttagagtgccatgcgaatgcttgaattgggtgtggagctcgat 1620
DB 1860 gaaactgactttagagtgccatgcgaatgcttgaattgggtgtggagctcgat 1919
QY 1621 tccctcgaagtgtcgaacattataaagatggaagccctattgcagacgacttgcgtt 1680
DB 1920 tccctcgaagtgtcgaacattataaagatggaagccctattgcagacgacttgcgtt 1979
QY 1681 gaaaaacttgttcttagtgctgccaacacaggtcttctgaacctgcgccagattgtt 1740
DB 1980 gaaaaacttgttcttagtgctgccaacacaggtcttctgaacctgcgccagattgtt 2039
QY 1741 ttgagcaaatgtgacagctctcttcataccaacacacatcgtggtgacgtgaatgat 1800
DB 2040 ttgagcaaatgtgacagctctcttcataccaacacacatcgtggtgacgtgaatgat 2099
QY 1801 gccaaatactgctcagaaatattagagttgcagctacttccaggcacaataatgccagct 1860
DB 2100 gccaaatactgctcagaaatattagagttgcagctacttccaggcacaataatgccagct 2159
QY 1861 acctttggacatttgcagggggatacagatggccaattattgatatcttttggagtga 1920
DB 2160 acctttggacatttgcagggggatacagatggccaattattgatatcttttggagtga 2219
QY 1921 gtattttccatggatatgtttttacagctgtttttaaaaaagagggataatgaatccagag 1980
DB 2220 gtattttccatggatatgtttttacagctgtttttaaaaaagagggataatgaatccagag 2279
QY 1981 gttggaatgaatacagaacaaactaatctgaaacctgggggatctctgacggcattggac 2040
DB 2280 gttggaatgaatacagaacaaactaatctgaaacctgggggatctctgacggcattggac 2339
QY 2041 atgctccacaatttctgaaacgtgagccaaacccaaagcgttcttaataatgagttaggc 2100
DB 2340 atgctccacaatttctgaaacgtgagccaaacccaaagcgttcttaataatgagttaggc 2399
QY 2101 ctgcatgctcyctga 2115
DB 2400 ctgcatgctccgtga 2414

RESULT 2
ID AAQ27307
XX AAQ27307 standard; cDNA; 2148 BP.
AC AAQ27307;
XX AAQ27307;
DT 03-FEB-1993 (first entry)
XX
DE Rabbit endopeptidase cDNA (preliminary sequence).
XX
KW microsome; metalloproteinase; polymerase chain reaction; ss.
XX
OS Oryctolagus cuniculus.
XX
FH Key Location/Qualifiers
CDS 11..2137
FT /tag= a
FT /product= endopeptidase
FT /note= "the identity of many of the nucleotides
is undefined"
XX
PN W09213080-A.
XX
PD 06-AUG-1992.
XX
PF 28-JAN-1992; 92WO-US00731.
XX

```

```

PR 28-JAN-1991; 91US-0646997.
XX (UNIW ) UNIV WASHINGTON.
PI Davie BW, Kawabata S;
XX WPI: 1992-284665/34.
DR P-PSDB; AAR26114.
XX
PT Microsomal endo-peptidase isolated from rabbit liver - used to
PT cleave protein substrates in-vivo and in-vitro
XX
PS Example 3; Fig 5; 48pp; English.
XX
CC The rabbit endopeptidase was purified from rabbit liver and
CC digested with CNBr. Families of degenerate primers were designed
CC based on the amino acid sequences of the CNBr fragments. The
CC primers were used to amplify sequences from a rabbit cDNA library.
CC An amplified fragment was isolated and used to probe the same
CC library. Two overlapping clones produced a 2148bp sequence. A
CC preliminary sequence analysis was performed in which much of the
CC 5' and 3'-most sequences are unassigned; the actual amino and
CC carboxyl terminals of the protein have not yet been determined.
CC The short untranslated 5' and 3' sequences are primer sequences.
CC See AAR26107-826113 and AAQ27305-Q27307.
XX
SQ Sequence 2148 BP; 568 A; 452 C; 551 G; 506 T; 71 other;

```

```

Query Match 75.2%; Score 1590.8; DB 13; Length 2148;
Best Local Similarity 87.2%; Pred. No. 0;
Matches 1728; Conservative 1; Mismatches 247; Indels 6; Gaps 1;

QY 1 atgategcccggtgccttttggctgtgcgaagcctccgcagagttgtgtgtccaggatt 60
DB 167 atnategcccggtgcttttccgctgtncgaggcctccacagggtgtgtgtccaggatt 226
QY 61 ttactcagaatcagcttagaagaagaagtgatgtctcttcacgacaatgtcttctctat 120
DB 227 ctattcaaatgacatcaggagaagaagtgatgtcaccctcttcacgacgtttcttctat 286
QY 121 actgtgctggcagaaaatgttttaagatgggatctttccaccagagcaaatataaaacaaga 180
DB 287 actgcagctggcagaatgttttaagatgggatctttccaccagagcaaatcaagaacaaga 346
QY 181 actgaggagctcattgtgcagaccacaaacgggtgtacgtgtgttggaaatgctcgatt 240
DB 347 actgaggagctcattgtgcagaccacaaacgggtgtatgtattctgttggaaatgctggatc 406
QY 241 gaggaagttaacttacgagaaactgtctgcagcactgtgcagatgtagaagttaagtata 300
DB 407 aaggagctgacttaacgagaaactgtctgcaggcgtgtgcaggcgtgtggaagtgaagtacata 466
QY 301 gtggaaagagaccatgctagactttcccccagcatgtatcctctgacaaaagaatgacagca 360
DB 467 gtggaaagaacctgctagactttctctcagcatgtttctactgacagagaagtacaggca 526
QY 361 gcaagtacagaagcagacaaaagactttctcgtttttgattattgagatgagcatgagagga 420
DB 527 gcaagtacagaagcagacaaagggctttctcgtttttgattattgagatgagcatgagagga 586
QY 421 gatattttgagagaattgttcatattacaggaacactgtgactcgtgggaagataaaacct 480
DB 587 gatattttcagagaatcgttccactacaggaaacactgtgactcgtgggaagataaaacct 646
QY 481 gagggcagacgactcttgaaagtcaattataatggggaagaagaatgggtcccatctt 540
DB 647 gaagccagacgatacttggaaagtgcagttaaaatggggaagaaatgggtcccatctt 706
QY 541 cctgaaacagtagcagaatgaaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600
DB 707 ccnnagaagtagcagaatganatcaaatcaatgaagaaagaatgagtgagctatgtatt 766

```

Qy	601	gattttaacaaaaaccctcaatgagatgatgaactcttccttgattatttccaggctgaactt	660
Db	767	gactttaacaaaaaccctcaatnaggatgataccttccntgtgttttccaggctgagctt	826
Qy	661	ggtagctctccctgagtgatttcattgacaggtttagaaaaagacagatgatgacaagtataa	720
Db	827	ggtagctctccctgagtgatttcattgacaggtttagaaaaagatggagatgacaaagtataa	886
Qy	721	attaccttaaaataccacactatttccctgctcatgaagaataattgtatccctgaaacc	780
Db	887	attaccttaaaataccacactatttccctgctcatgaagaataattgtattctctgaaacc	946
Qy	781	agaagaagatgaaatggcttttaatacaagggtcaaaaggagaaacacacataatttg	840
Db	947	agaaggagaatggaatggcttttaatacaagatgcaaaaggagaaacacgcctattctg	1006
Qy	841	caagcagtactcccactgcgaaccgaaggtggccaaactactcgttatagcacacatgct	900
Db	1007	cagcagctactcccactgcggcccaagtggccaaactcctgggctatagcacacatgct	1066
Qy	901	gacttcgtcttgaatgaacactgcaagagacgaacgcgcgtacacagcgtttcttagat	960
Db	1067	gacttcgtcttcgagatgaacactgcaagagacgaacgcgcgtacacagcgtttcttagat	1126
Qy	961	gatttaagccagaagttaaaccccttgggtgaagcagaacgagaggttatttgaatttg	1020
Db	1127	gatttaagccagaataataaacgcgtgggtgaggcagaacgagaggttattcttgagttg	1186
Qy	1021	aagaaaaaggaatgcaagacacaggggttttgaatatgatgggaaaaatcaatgcctgggat	1080
Db	1187	aagaaaaaggaatgcgaagagaaaggtttgagtatgacgggaaaaatcaatgcctgggac	1246
Qy	1081	ctatatctacatgaactcagacagagaactcgaattatccatagaccagaaggttcctc	1140
Db	1247	ctcatctactacatgactcagacagagaactcgaattatccatagaccagaaggttcctc	1306
Qy	1141	aaggaaatacttcccgaattgaggtgtgcactgaa-----ggctgtcgaacacctaccag	1194
Db	1307	aaggagtaacttccccccnntnagggttgctcancnaggggcctgctgaaactctaccag	1366
Qy	1195	gagttgtgggacttttcatttgaaacaaatgacagatgctcatgttttggaaacaagagtgtt	1254
Db	1367	gagttgtgggacttttattcgagcaagtggccgacgtcatgttttggaaaccgaagntc	1426
Qy	1255	acactttacttgtaaggataaagctacagagagaagtattgggacaggtttctatttggac	1314
Db	1427	acactttacacttgtaaggacaaaggtctacagaggaagtgcgtggggcaggttctaccctggac	1486
Qy	1315	ctctatccaaaggaaagaaaaatacaatcatcgccctgcttccgctccagcctgctgc	1374
Db	1487	ctctatccngggaaagaaaaataaatcatgcagncigtcttggctccagcctggctgc	1546
Qy	1375	ctctctgcttgatggaaagccggatgatggcgtgcctctgcgttggatgaaactctcacag	1434
Db	1547	ctgctncccgatgggacggatgctgctgtggtgcctctggttgaactnnnccag	1606
Qy	1435	ccagtggcaggtcgctccctctctctgagacacagcagaggtgaggacctactttcatgag	1494
Db	1607	ccagtggcggccgcnccctctctctgagggcgatgaagtgmgnctctacnccccatgan	1666
Qy	1495	tttgggtcacgtgatgcatcagatttgcacagactgatttgcagcatattagcgaaca	1554
Db	1667	tttgggtcatgcatcagctttgtgcnacagactgatttgcagcatattagtgaaaca	1726
Qy	1555	aatgtggaacactgactttgtagagdgccatcgcaaatgcttgaaaaattgggtgtggggac	1614
Db	1727	aatgtggaacactgactttgtagagdgcccttcaaaatgcttgagaactgggtgtgggac	1786
Qy	1615	gtcgattccctccgaagattgtcaaaacattataaagatgaagccctattgcagacgat	1674
Db	1787	atcgactctctccgangaatnnmaaacatataaagatggnnaccctattgcagatgat	1846
Qy	1675	ctgcttggaaaaactgtgtcttcttagctgtgtgcacacagaggtcttctgcacccctgcgccag	1734

Db	1847	ctctctgaaancttgtctctctagactggtcaacacaggctctctgaccccttcgccaag	1906
Qy	1735	attgttttgagcaaaagtgtgatcagtcctcttcataccacacatcgtggtgctgcaagt	1794
Db	1907	atcgttttgagcaaaagttagccagtcctctgcataccaactcatcctggacgctgcaagt	1966
Qy	1795	gaatatgccaaatactctcagaataattaggaggttcagctactctccaggacacaaatatg	1854
Db	1967	gaatatgccaggtactgacccgacattttagtggtgcgtacnccctggaaacacatg	2026
Qy	1855	ccagctacacctttggacatttggcagggggatcagatggccaatatattggatatctttgg	1914
Db	2027	ccagctacttttggcatttggcannnggatcagatgccagttatttgatatctttgg	2086
Qy	1915	agtgaaagtatttccatcgatatgtttttacagctgttttaaaaaagaaggataatgaat	1974
Db	2087	agtgaaagntttccatggacatgtttntacagctgttttaaaaaaaaanaacccggaat	2146
Qy	1975	cc 1976	
Db	2147	tc 2148	
RESULT 3			
AAH14379			
ID	AAH14379 standard; cDNA; 1658 BP.		
XX			
AC	AAH14379;		
XX			
DT	26-JUN-2001 (first entry)		
XX			
DE	Human cDNA sequence SEQ ID NO:11796.		
XX			
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss		
XX			
OS	Homo sapiens.		
XX			
PN	EP1074617-A2.		
XX			
PD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-0116126.		
XX			
PR	29-JUL-1999; 99JP-0248036.		
XX			
PR	27-AUG-1999; 99JP-0300253.		
XX			
PR	11-JAN-2000; 2000JP-0118776.		
XX			
PR	02-MAY-2000; 2000JP-0183767.		
XX			
PR	09-JUN-2000; 2000JP-0241899.		
XX			
PA	(HELI-) HELIX RES INST.		
XX			
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
DR	WPI: 2001-318749/34		





DR P-PSDB; AAR24259.

XX Human amyloidin protease - used for cleaving Met-Asp bond in  
PT amyloid-like substrate for identifying protease inhibitors

XX Claim 37; Page 53-54; 62pp: English.

XX The complete sequence of the human amyloidin gene was obtained by  
CC sequencing two clones: clone CHL57 (codons 1-480) and clone CHL53  
CC (codon 56 to STOP codon). The clones were isolated after screening cDNA  
CC libraries prepared from the following sources: normal human brain,  
CC superior temporal gyrus and human embryonic kidney cell line 293.  
CC Examination of the nucleotide sequence reveals that there is no  
CC clear hydrophobic leader sequence. The deduced amino acid sequence  
CC of the amyloidin protease contains the motif X-H-E-F-G-H-X at the  
CC active site typical of zinc-dependent metalloproteinases; the two  
CC histidine residues coordinate the Zn (2+) in the active centre and  
CC the glutamate is involved in bond-breaking.  
CC See also AAR24260-3, AAR24266-7 and AAQ24876-Q24887.

XX Sequence 2070 BP; 434 A; 626 C; 578 G; 332 T; 0 other;

Query Match 37.4%; Score 790.8; DB 13; Length 2070;  
Best Local Similarity 62.4%; Pred. No. 3.7e-219;  
Matches 1236; Conservative 1; Mismatches 743; Indels 0; Gaps 0;

QY 128 ctggcagaaatgttttaagatgggattcttccacagagcaaatataaaacagaactgagg 187  
DB 56 ctgttgtaaacgactgcgtgggacctgagtcgccacagatagaggagcgccacagg 115  
QY 188 agctcattgtgcagacaaacagggtgtacgatgctgttggaatgcctggtattgaggag 247  
DB 116 agctcatgcagacaaacaggcggtgatgaccagggttggcaccaggagttgaggagc 175  
QY 248 taacttacgagaactgtctgcaggcaactggtcagatgtagaagtaaatatagtgga 307  
DB 176 tgcctacagagacgcgtcgaaggcggtgcgatgtgaggtgcactacacagttcaga 235  
QY 308 ggacctctagactttcccagactgtatctctgcacaaagagtagcagcaaga 367  
DB 236 ggaatattcttgacttcccagactgttcccctccaaaggacatccggacacagca 295  
QY 368 cagaagcagacaaagactttctgtttgatattgagatgagatgagagagatatat 427  
DB 296 cagaggccgacaaagactctctgagttcgcgtggagatgagatgagagagcgtgt 355  
QY 428 ttgagagaattgttcattacaggaacactgtgactgtgggaagataaaacctgaggcca 487  
DB 356 accagagagatcgtgtgctccaggagaaagtctcagaaggactcactgaggccgagcgtg 415  
QY 488 gacgatacttgaaaaagtcattaaatgggaaaaagaaatgggctccacttcttgaa 547  
DB 416 cgcgttacttgagcggcgtatcaagctggccggcgggaaatgggttccactcccccag 475  
QY 548 aagtcagaaatgaatacaatcaatgaagaaagagtagtgagctatgtattgatttta 607  
DB 476 agactcagaaaaacatacaacgcatacaagaagaagctgagccttctgtgcatgcattca 535  
QY 608 acaaaaacctcaatgaggatgataccttcccttgtattttccaaggctgaaattggtgctc 667  
DB 536 caaagaacctgaacgaggacacgaccttccctgcccctcacgctccacgagctaggaggc 595  
QY 668 ttctgatgatttcattgacagtttagaaaaagacagatgatgacaagtataaaattact 727  
DB 596 tcccogaggactttctgaactccctcctggagagatggaggacggaagttgaggtcacccc 655  
QY 728 taaataaccacactatttccctgctcatgaagaaatgttgtatccctgaaacccagaagaa 787  
DB 656 tcaagtacccttacttccctcctctgaagaaatgccacgtcctgagaccaggagga 715  
QY 788 ggaatggaatggcttttaatacagaaggtgcaaggaggaacacccataattttgcagcgc 847

DB 716 aagtggaggaggccttcaactgccggtgcaaggaggagaactgcgtatctctcaaggagc 775  
QY 848 tactcccactgcgaacaaagtgccaaactactcgggttatagcacacatgctgacttcg 907  
DB 776 tggcgacgtcggggccaggaagtcocgcctcgtcgtgggtttccacacgacccgactatg 835  
QY 908 tccttgaatgaacactgcgaagacaaagcgcgttaacagcgttttctagatgatttaa 967  
DB 836 tcctggagatgaactggcgaagcagaccgagaccgtggccaccttctagatgagctgg 895  
QY 968 gccgaagtttaaacaccttgggtgaagcagaaacagagatttatttgaattgagaanaa 1027  
DB 896 cgcgaagctgaagccctggggagcagcgctgcgtgattcttggagctgaagcgtg 955  
QY 1028 aggaatcaaaagacaggggttttgaatatgatggaaaatcaatgcctgggatctatt 1087  
DB 956 cggagtgcgagcggcggggctgccttcgacgcccgcctcgtgcttgggacatgcgct 1015  
QY 1088 actacatgactcagacagaggaactcaagttatccatagacacaaagagttcctcaaggaa 1147  
DB 1016 actacatgaaccaggtggaggagacgcgctactcgtggaccgaacactgctcaaggag 1075  
QY 1148 acttccaaattgaggtgctactgaagccttgcgtgaacacactaccagagattgttgggac 1207  
DB 1076 acttcccctgtaggtgctacgcacggcgtcgtgggcatactaccagagactcctggggc 1135  
QY 1208 ttctattgaacaaatgacagatgctcatgttttgaaacaaagagtttacctttatactg 1267  
DB 1136 tggccttcacacagagagggcgccagtcgctggcagtgagagagtgcggtctacacgg 1195  
QY 1268 tgaaggataaagctacagggagaagtatttggacagtttctatttggacctctatccaaagg 1327  
DB 1196 cgaggcagcgcgctcggggagggtggtcggaagtcttactctgacctgtaccgcggg 1255  
QY 1328 aagaaaaatacaatcatcgtcggcctgcttcgcacccgtgccttctgctgctgatg 1387  
DB 1256 aagaaaaatgcggcagcggcctgcttggcgtgcagcccggcgtcgtcggcgaggtg 1315  
QY 1388 gaagccggatgatggcagtggtgcctcgtggtgaaacttctcagaccagtggtggagctc 1447  
DB 1316 ggagcccgagatgcctacgcggccatggtggccatggtggccaaacttcaccaagcccacgagc 1375  
QY 1448 gtccctctctctgagacacagcaggtgaggaacttacttctcatgatttggctcaagcga 1507  
DB 1376 cgccctcgtcgtcagcatgacgaggtggagacctacttccatgagtttggcactga 1435  
QY 1508 tgcacagatttgcacagactaattttgcagatttgcaggaacaaatgtggaaactg 1567  
DB 1436 tgcaccagctcgtcccagcgaggttcgcacgttgcagcgggacccacgtggagcggg 1495  
QY 1568 actttagaggtgccatcgcaaatgcttgaataatgggtgtgggacgtcgtattccctcc 1627  
DB 1496 actttagagggccgctcgagatgctggagaactgggtgtgggacgagcgcgctgc 1555  
QY 1628 gaagattgtcaaaacattataaagatggaagccctatttgcagacgatctgttgaanaac 1687  
DB 1556 tgcggatgtcggcaactaccgcacagcagcgccgtgccccgggagctccttggagaagc 1615  
QY 1688 ttgttctttagctgttcaacacaggtcttctgacctgcgcagatttgggttggagca 1747  
DB 1616 tcatgtgctccggcagggccacacagcctcttcagcctgcgcagatcgtcctcgcca 1675  
QY 1748 aagtgtgactctcttccataccaacacatcgtggatgtgcaagtgaatatgccaat 1807  
DB 1676 aggtggaccagccctgcacgcagacgagcagaccccgcgaggtatgcgcggc 1735  
QY 1808 actgctcagaaaattattaggagtttgcagctactccaggcacaataatgcccagctactttg 1867  
DB 1736 tctgcaggagatcctcgtgggtcccgccagcggaacacacactgctgcaaccttcg 1795  
QY 1868 gacatttggcaggggatacagatggcaaatatttgatatcttttggagtgaaatattt 1927  
DB 1796 gccatctggcagggtgggtacagccccagttactcgggtacactgtggagcgaggcgtatt 1855



```
PD XX 06-APR-2000.
PF XX 23-SEP-1999; 99WO-US22226.
XX XX 28-SEP-1998; 98US-0102161.
PR XX 28-SEP-1998; 98US-0102180.
PR XX 29-SEP-1998; 98US-0102380.
PR XX 08-OCT-1998; 98US-0103815.
PR XX 27-OCT-1998; 98US-0105877.
XX XX (CHIR ), CHIRON CORP.
PA XX (HYSE-) HYSEQ INC.
XX XX Williams LT, Escobedo J, Inniss MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac K, Ranzan-Jakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX XX WPI; 2000-293155/25.
DR XX Polynucleotide library comprising 1079 defined sequences, useful in
PA the form of an array to detect cancer or susceptibility to cancer -
PT Claim 1; Page 464; 502pp; English.
XX XX The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (1); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Kml2L4-A CDNA library.
XX XX Sequence 339 BP; 97 A; 72 C; 81 G; 89 T; 0 other;

Query Match 16.0%; Score 339; DB 21; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.2e-88;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1548 cggacaataatgtgaaactgactttgttagagtgccatcgcaaatgcttgaattgggt 1607
Db 1 cggacaataatgtgaaactgactttgttagagtgccatcgcaaatgcttgaattgggt 60
Qy 1608 gtggagctgctatccctccgaagattgtcaaaacattataaagatggaagccctattgc 1667
Db 61 gtggagctgctatccctccgaagattgtcaaaacattataaagatggaagccctattgc 120
Qy 1668 agacgactgtgtgaaaaactttgttcttaggtgtgctgacacaggtctcttgaccct 1727
Db 121 agacgactgtgtgaaaaactttgttcttaggtgtgctgacacaggtctcttgaccct 180
Qy 1728 gcgccagattgtttgagcaaatgtgatcagctctcttccatcccaacacatcgctggtgc 1787
Db 181 gcgccagattgtttgagcaaatgtgatcagctctcttccatcccaacacatcgctggtgc 240
Qy 1788 tgcagtgaaatgccaataactgctcagaataattagagttgagctactccaggcac 1847
Db 241 tgcagtgaaatgccaataactgctcagaataattagagttgagctactccaggcac 300
Qy 1848 aaatgcccagctacacctttggacatttggcagggggata 1886
Db 301 aaatgcccagctacacctttggacatttggcagggggata 339
```

```
RESULT 7
AAC98797
ID AAC98797 standard; cDNA; 900 BP.
XX AC AAC98797;
XX DT 09-MAR-2001 (first entry)
XX DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:25.
XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
XX KW detection; diagnosis; identification; cytostatic; neuroprotective;
XX KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
XX KW antiinflammatory; cardiant; gene therapy; chromosome mapping;
XX KW linkage analysis; tissue identification; tissue typing; forensic;
XX KW neural; immune system; muscular; reproductive; gastrointestinal;
XX KW pulmonary; cardiovascular; renal; proliferative; ss.
XX OS Homo sapiens.
XX PN WO200055320-A1.
XX PD 21-SEP-2000.
XX PF 08-MAR-2000; 2000WO-US05989.
XX PR 12-MAR-1999; 99US-0124270.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2000-579444/54.
XX DR P-PSDB; AAB54032.
XX PT New nucleic acid that is a pancreatic cancer antigen for preventing,
XX PT treating, or ameliorating a medical condition, particular pancreatic
XX PT cancer, or for use in assays for diagnosing a pathological condition -
XX FS Claim 1; Page 512-513; 1379pp; English.
XX CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
XX CC proteins, called pancreatic cancer antigens, given in AAB54008 to
XX CC AAB54466. The human pancreatic cancer antigens have cytostatic,
XX CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
XX CC gynaecological, cardiant and antiinflammatory activities, and can be used
XX CC in gene therapy. The polynucleotide and proteins can be used for
XX CC preventing, treating, or ameliorating a medical condition or in assays
XX CC for diagnosing a pathological condition or a susceptibility to one in a
XX CC subject. Binding partners to the proteins and the activity of the
XX CC proteins can be identified. The pancreatic cancer antigens can be used to
XX CC detect, treat or prevent pancreatic disorders, especially cancer.
XX CC Agonists and antagonists to the antigens can be screened for. The
XX CC pancreatic cancer antigen polynucleotides can be used to design nucleic
XX CC acid hybridisation probes that can be used in chromosome mapping, linkage
XX CC analysis, tissue identification and/or typing and a variety of forensic
XX CC methods. The proteins can be used to generate antibodies
XX CC which are used to purify, detect and target the polypeptides. The
XX CC both in vivo and in vitro diagnostic and therapeutic methods. The
XX CC proteins can be used to treat or prevent neural, immune system, muscular,
XX CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
XX CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
XX CC sequences used in the exemplification of the present invention.
XX SQ Sequence 900 BP; 160 A; 284 C; 299 G; 147 T; 10 other;

Query Match 14.6%; Score 308.8; DB 21; Length 900;
Best Local Similarity 64.1%; Pred. No. 2.3e-79;
Matches 480; Conservative 0; Mismatches 268; Indels 1; Gaps 1;
```





```

Db 601 www..... 660
QY 1041 cagggttttgatgatgggaataatcaatgcctggggtatctattactacatgactca 1100
Db 661 www..... 720
QY 1101 gacagaggaactcaagtattccatagaccacagagttcctcaaggaatacttccaatt 1158
Db 721 www..... 778

RESULT 10
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 5.3%; Score 112.8; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 2e-22;
Matches 6; Conservative 475; Mismatches 297; Indels 0; Gaps 0;

QY 381 aagactttctgttttgatgatgagcagcagagagagatatatttgagagaattgt 440
Db 1 www..... 60
QY 441 tcattacaggaacacctgtatctggggaagataaaacctgagccagacgatacttga 500
Db 61 www..... 120
QY 501 aaagtcaattaaatggggaagaaatggggtccatcttctgaacaaagtacagaatga 560
Db 121 www..... 180

```

```

QY 561 aatcaaatcaatgaagaagaatgagtgagctatgtattgatttaacaaaacctcaa 620
Db 181 www..... 240
QY 621 tgaggatgatacttctgttatttccaaggctgaacttggctcttctctgatgttt 680
Db 241 www..... 300
QY 681 cattgacagtttagaaaaagacagatgatgacaagtataaaattacattcaaaatccaca 740
Db 301 www..... 360
QY 741 ctatttccctgctcatgaagaatgttgatccctgaaacacagagaagatggaatggc 800
Db 361 www..... 420
QY 801 tttaatacaaggtgcaagagagaaacacacataattttgcagcagactactccactgcg 860
Db 421 www..... 480
QY 861 aaccaaggtggccaaactactcgtttatagcacacatgctgacttgccttgaaatgaa 920
Db 481 www..... 540
QY 921 cactgcaaaagagcacaaagccgggttaacagcytttctagatgatttaagccagaagttaa 980
Db 541 www..... 600
QY 981 acccttgggtgaagcagaacgagagtttattttgaatttgaaagaaaaggaagcaaga 1040
Db 601 www..... 660
QY 1041 cagggttttgatgatgggaataatcaatgcctggggtatctattactacatgactca 1100
Db 661 www..... 720
QY 1101 gacagaggaactcaagtattccatagaccaagagttcctcaaggaatactccaatt 1158
Db 721 www..... 778

```

```

RESULT 11
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

```











Search completed: January 14, 2002, 20:49:53  
Job time: 6452 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 14, 2002, 18:38:36 ; Search time 72.17 Seconds

(without alignments)  
722.566 Million cell updates/sec

Title: us-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCILLAVSLRRVGGSR.....FLKREPQKAFLMRGLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

A.Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3668	100.0	704	22 AAB73485	Human aminopeptidase
2	2979	81.2	709	13 AAR26114	Rabbit endopeptidase
3	2331	63.5	689	13 AAR24259	Human amyloidin pr
4	1381	54.0	381	22 AAB93028	Human protein sequ
5	702	19.1	678	21 AAB58575	N. gonorrhoeae amin
6	702	19.1	678	21 AAB58577	N. meningitidis am
7	702	19.1	678	21 AAB58577	N. gonorrhoeae amin
8	702	19.1	678	21 AAB58577	N. meningitidis am
9	702	19.1	678	21 AAB58577	N. meningitidis am
10	702	19.1	678	21 AAB58577	Neisseria gonorrhe
11	702	19.1	678	21 AAB58577	Neisseria meningit

12	702	19.1	678	21	AA75802	Neisseria meningitidis am
13	694	18.9	678	21	AAB38576	N. meningitidis am
14	694	18.9	678	21	AAB38579	N. meningitidis am
15	694	18.9	678	21	AAB25646	N. meningitidis am
16	694	18.9	678	21	AA74506	Neisseria meningitidis am
17	694	18.9	678	21	AA74509	Neisseria meningitidis am
18	694	18.9	678	21	AA75801	Neisseria meningitidis am
19	694	18.9	678	21	AA75804	Neisseria meningitidis am
20	691	18.8	678	21	AAB25649	N. meningitidis am
21	659.5	18.0	783	21	AAG51467	Arabidopsis thaliana
22	659.5	18.0	785	21	AAG51466	Arabidopsis thaliana
23	659.5	18.0	793	21	AAG51465	Arabidopsis thaliana
24	583	15.9	678	22	AAG92271	C. glutamicum prote
25	567	15.5	211	21	AAB34032	Human pancreatic c
26	480.5	13.1	476	21	AAB38574	N. meningitidis am
27	480.5	13.1	476	21	AAB25644	N. meningitidis am
28	475	12.9	491	21	AAB38578	N. gonorrhoeae amin
29	475	12.9	491	21	AAB25648	N. gonorrhoeae amin
30	475	12.9	491	21	AA74507	Neisseria gonorrhoeae
31	475	12.9	491	21	AA75803	Neisseria gonorrhoeae
32	461.5	12.6	344	21	AA74505	Neisseria meningitidis
33	461.5	12.6	344	21	AA75799	Neisseria meningitidis
34	446.5	12.2	555	22	AA79072	Corynebacterium gl
35	381	10.4	355	22	AA79073	Corynebacterium gl
36	270	7.4	234	11	AA04011	Protein product of
37	169	4.6	41	22	AA18211	Peptide #4645 enco
38	169	4.6	41	22	AA30705	Peptide #4742 enco
39	169	4.6	41	22	AA05830	Peptide #4512 enco
40	146	4.0	29	21	AA53253	Human type enzyme
41	133	3.6	1392	20	AA06999	Restin protein seq
42	132.5	3.6	1294	22	AA41110	Human polypeptide
43	131	3.6	1427	12	AA10534	Human 160kD mediat
44	128.5	3.5	2048	22	AA40027	Human polypeptide
45	127.5	3.5	1700	22	AA39324	Human polypeptide

## ALIGNMENTS

RESULT	1
AA73485	
ID	AA73485 standard; Protein; 704 AA.
XX	
AC	AA73485;
XX	
DT	25-JUL-2001 (first entry)
XX	
DE	Human aminopeptidase 22196.
XX	
KW	Human; aminopeptidase 22196; neutral zinc metalloproteinase;
KW	metalloenzyme; analgesic; cytostatic; lung cancer; colon cancer; tumour;
KW	pain; drug screening; spleen disorder; infectious disease;
KW	immune disorder; lung disorder; colon disorder; liver disorder;
KW	uterine disorder; brain disorder; skin disorder; cardiac disorder;
KW	vascular disorder; bone disorder; gene therapy.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	Modified-site 11..13
FT	/note= "Ser is O-phosphorylated by protein kinase C"
FT	Modified-site 59..62
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 104..107
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 114..116
FT	/note= "Ser can be O-phosphorylated by protein kinase C or by casein kinase II"
FT	Modified-site 123..126
FT	/note= "Thr is O-phosphorylated by casein kinase II"
FT	Modified-site 127..130
FT	/note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"
FT	

FT Modified-site 130..133 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 137..139 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 157..165 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 169..171 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 172..175 /note= "Amidation site"  
FT Modified-site 190..192 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 193..196 /note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"  
FT Modified-site 216..219 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 233..239 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 234..237 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 242..244 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 260..262 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 298..301 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 308..310 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 312..314 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 323..325 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 366..369 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 392..397 /note= "Gly is N-myristoylated"  
FT Modified-site 396..399 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 415..418 /note= "Asn is N-glycosylated"  
FT Modified-site 422..425 /note= "Thr can be O-phosphorylated by protein kinase C or by casein kinase II"  
FT Modified-site 453..458 /note= "Gly is N-myristoylated"  
FT Modified-site 475..478 /note= "Asn is N-glycosylated"  
FT Modified-site 488..495 /note= "Tyr is O-phosphorylated by a tyrosine kinase"  
FT Modified-site 518..521 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 541..543 /note= "Ser is O-phosphorylated by protein kinase C"  
FT Modified-site 543..546 /note= "Ser is O-phosphorylated by cAMP/cGMP-dependent protein kinase"  
FT Modified-site 552..557 /note= "Gly is N-myristoylated"  
FT Modified-site 575..577 /note= "Thr is O-phosphorylated by protein kinase C"  
FT Modified-site 582..585 /note= "Ser is O-phosphorylated by casein kinase II"  
FT Modified-site 591..594 /note= "Asn is N-glycosylated"  
FT Modified-site 592..595 /note= "Thr is O-phosphorylated by casein kinase II"  
FT Modified-site 627..632 /note= "Gly is N-myristoylated"  
FT Modified-site 674..679 /note= "Gly is N-myristoylated"  
XX PN WO200123590-A2.

XX 05-APR-2001.  
XX PD  
XX PF  
XX PR 02-OCT-2000; 2000WO-US27214.  
XX PR 30-SEP-1999; 99US-0409180.  
XX PA (MILL)-MILLENNIUM PHARM INC.  
XX PI Kapeller-Libermann R, White D, Silos-Santiago I;  
XX DR WPI; 2001-300101/31.  
XX DR N-PSDB; AAF84499.  
XX PT Novel human aminopeptidase, used to treat colon and lung cancer, and to  
XX PT identify modulators used to treat e.g. cirrhosis  
XX PS Claim 1; Fig 1A-B; 109pp; English.  
XX  
CC This sequence represents a novel human aminopeptidase, designated  
CC 22196. cDNA clones encoding human aminopeptidase 22196  
CC were identified in a bone marrow cDNA library using an EST (expressed  
CC sequence tag) selected on the basis of homology to aminopeptidase  
CC sequences, and were assembled to give the full-length cDNA.  
CC Aminopeptidase 22196 is thought to be a metalloenzyme, as it contains  
CC a zinc-binding region signature found in neutral zinc metalloproteinases.  
CC Expression analysis indicated that the aminopeptidase is highly  
CC expressed in normal osteoblasts, testis, skeletal muscle, foetal kidney  
CC and foetal liver. It is also expressed in normal breast, lung, and colon  
CC tissue, and is overexpressed in lung and colon cancer. Expression was  
CC found to be downregulated during stromal cell osteoblast lineage  
CC maturation. Aminopeptidase 22196 may be used to treat pain, and cancers  
CC of the lung and colon. It can also be used in drug screening to identify  
CC compounds which bind to it and modulate its activity. Such compounds may  
CC be used in the treatment of a wide variety of conditions, including  
CC spleen disorders (e.g., splenomegaly); infectious diseases (e.g.,  
CC malaria, tuberculosis); immune disorders (e.g., rheumatoid arthritis,  
CC systemic lupus erythematosus, acquired immunodeficiency syndrome (AIDS));  
CC lung disorders (e.g., pulmonary embolism, emphysema, bronchial asthma);  
CC colon disorders (e.g., diarrhoea, Crohn's disease, ulcerative colitis);  
CC liver disorders (e.g., hepatic cirrhosis, hepatitis); uterine disorders  
CC (e.g., endometriosis); brain disorders (e.g., meningitis, Alzheimer's  
CC disease, Parkinson's disease, Huntington's disease); skin disorders  
CC (e.g., vitiligo, urticaria, skin tumours, psoriasis); cardiac and  
CC vascular disorders (e.g., myocardial infarction, chronic ischaemic heart  
CC disease, hypertension, atherosclerosis); and bone disorders (e.g.,  
CC osteoporosis, Paget's disease). Nucleotides encoding aminopeptidase 22196  
CC may be used in gene therapy.  
XX  
SQ Sequence 704 AA;  
  
Query Match 100.0%; Score 3668; DB 22; Length 704;  
Best Local Similarity 100.0%; Pred. No. 9.5e-308;  
Matches 704; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MIARCLAVRSLRVRGSRILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR 60  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 TEELIVQTKQYDVGMLGTEVTYENCQLQALADVEVKYIVERTMLDPQHVSSDKVEYA 120  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 61 teelivqtkvydvgmlgteevtyencqlqaladvevkyyivertmldfpqhvsdkveya 120  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 121 ASTEADKRLSRFDTEMGRGDIFERIVHLQETDGLGKIKPEARYLEKSIKMGKRNGLHL 180  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 PEQVQNETKSMKRWSELCDIDFNKLNEDDFLVFSKAEALGALPDDFDSLEKTDGDKYK 240  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 181 pegvqneiksmkrmseicidfnklnedddflvfskaelgalpddfdslsektdddkyk 240  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
QY 241 ITLKYPHYFPVNMKKCCIPETRRRMEAFNTRCKEENTIILOQLPLRLTKVAKLLGYSTHA 300

||||| itlkyhyfvmkkccipetrrrmemafrnckeeentilqlplrtkvakllystha 300  
Db 241  
QY 301 DFVLEMTAKSTSRVTAFLDLSOKLPLGEAREFELNKKKECKDRGFEDGKINAWD 360  
Db 301 dfvlemtakstsrvtafldlsqklplgeaerefilnkkeckdrgefvdgkinawd 360  
QY 361 LYYMTQTTEELKYSIDQELKEVFPFIEVTEGLLNTYQELLGLSPQMTDAHVWKNKSVTL 420  
Db 361 lyymtqtteelkysidqelkevfpiervteglntyqellglstfeqmtdahwnksvtl 420  
QY 421 YTVKDKATGEVLGGFYLDLYPREGKYNHAAACFGLOPGCLLPDGSRMMAVAALVYVNFSPQV 480  
Db 421 ytvkdkatgevlgfyldlypregkynhaacfglpgcllpdgsrmmaavaalvvnfsqpv 480  
QY 481 AGRPSLLRHDEVRTYFHFEGHVMHQICAQTDFAFSGTNVETDFVEVPSQMLENWWVDYD 540  
Db 481 agrpsllrhdevrtyfhfeghvmhqicactdfarfsgtnvetdfvevpsqmlenwwvdvd 540  
QY 541 SLRRLSKHYKDGSPYADDDLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASEY 600  
Db 541 slrrlskhykdgspiadddlleklvasrlvntgltlrqivlskvqdshtntsltdaasey 600  
QY 601 AKYCSILGVAATPGTNMPATFGHLAGGYDGYGYLYWSEVFSMDMFYSCFKKEGIMNPE 660  
Db 601 akycsilgvaatpgtnmpatfghlaggydgygylywsevfsmfyscfkkgimnpe 660  
QY 661 VGMKYRNLLKGGSLDGMDLNHLNFKRPNQKAFILMSRGLHAP 704  
Db 661 vgmkyrnllkpggsldgmdlnhlnfkrepnqkafilmsrglhap 704

## RESULT 2

AAR26114  
ID AAR26114 standard; Protein; 709 AA.  
XX AC AAR26114;  
DT 03-FEB-1993 (first entry)  
XX DE Rabbit endopeptidase (preliminary sequence).  
XX microsome; metalloproteinase; polymerase chain reaction.  
XX OS Oryctolagus cuniculus.  
XX PN WO9213080-A.  
XX PD 06-AUG-1992.  
XX PF 28-JAN-1992; 92WO-US00731.  
XX PR 28-JAN-1991; 91US-0646997.  
XX (UNIW ) UNIV WASHINGTON.  
XX PA Davie EW, Kawabata S;  
XX PI WPI; 1992-284665/34.  
XX DR N-PSDB; AAQ27307.  
XX Microsomal endo-peptidase isolated from rabbit liver - used to  
XX cleave protein substrates in-vivo and in-vitro  
XX Example 3; Fig 5; 48pp; English.  
XX The rabbit endopeptidase was purified from rabbit liver and  
XX digested with CNBr. Families of degenerate primers were designed  
XX based on the amino acid sequences of the CNBr fragments. The  
XX primers were used to amplify sequences from a rabbit cDNA library.  
XX An amplified fragment was isolated and used to probe the same  
XX library. Two overlapping clones produced a 2148bp sequence. A  
XX preliminary sequence analysis was performed in which much of the

CC 5' and 3'-most sequences are unassigned; the actual amino and  
CC carboxyl terminals of the protein have not yet been determined.  
CC The "Others" in the sequence correspond to codons in which 1 or more  
CC of the bases has not yet been identified.  
CC See AAR26107-R26113 and AAQ27305-Q27306.  
XX SQ Sequence 709 AA;  
Query Match 81.2%; Score 2979; DB 13; Length 709;  
Best Local Similarity 86.7%; Pred. No. 2.8e-248;  
Matches 568; Conservative 29; Mismatches 56; Indels 2; Gaps 1;  
QY 2 IARCLLAVSLRRVSGSRILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTRT 61  
Db 54 iarcfsaxrghrvsgsrillfkmrlgrvmsplqavssytaagnvrlwdlspqiktrt 113  
QY 62 EELIVOTKQYDAVGMGLGIEEVTYENCALQADVEYKIVERTMLDFPQHVSDDKEVRAA 121  
Db 114 eeliaqtqkydsvgmldikdvtynclqaladvekyivertmldfpqhvstdrevraa 173  
QY 122 STEADKRLSRFDIEMSRGDIPIFIVHLOETCDLGIKPEARRYLEKSIKMGKRNGLHLP 181  
Db 174 steadkrlsrfdiemsmredifqirivhlqetcdlekikpearryleksgvmgrxglhlx 233  
QY 182 EOYONEIKSMKRMSELCDIFNKNLNEDETFLVFSKAEALGALPDDFDLSLEKTDOKYKI 241  
Db 234 xevgnxiksmkrmseicidfnknlnxdddtxvfkaelgaipddfdidslekmdddkykl 293  
QY 242 TLKYPHYFPVMKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVAKLGYSTHAD 301  
Db 294 tlkyphyfvmkkccipetrrrmemafntrckexntvilqllplraqvakllygsthad 353  
QY 302 FVLEMTAKSTSRVTAFLDLSOKLPLGEAREFELNKKKECKDRGFEDGKINAWDL 361  
Db 354 fvlemtakstsrvtafldlsqklplgeaerefilnkkeceekgefeydginawdl 413  
QY 362 YXYMTQTEELKYSIDQELKEVFPFIEV--VTEGLLNTYQELLGLSPQMTDAHVWKNKSVT- 419  
Db 414 hyymtqteelkysidqelkefeypxxxgxxxrgllniyqellglxfeqvadanwnpxvt 473  
QY 420 LYTVDKATGEVLGGFYLDLYPREGKYNHAAACFGLOPGCLLPDGSRMMAVAALVYVNFSPQ 479  
Db 474 lytvdkatgevlgfyldlypexegkynhaxcfglpgclxpdgstrmlsvaalvvnxxqp 533  
QY 480 VAGRPSLLRHDEVRTYFHFEGHVMHQICAQTDFAFSGTNVETDFVEVPSQMLENWWVDV 539  
Db 534 vagrpsllrhdevrtyfhfeghvmhqicxqtdfarfsgtnvetdfvevpsqmlenwwvdi 593  
QY 540 DSLRRLSKHYKDGSPYADDDLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASE 599  
Db 594 dsrlrxxxkhykdxpiaddlleklvasrlvntgltlrqivlskvqdshtntsltdaase 653  
QY 600 YAKYCSILGVAATPGTNMPATFGHLAGGYDGYGYLYWSEVFSMDMFYSCFKKE 654  
Db 654 yaryctdilgvaaxpgtnmpatfghlaggydgygylywsexxsmdmxyscfkxxx 708

## RESULT 3

AAR24259  
ID AAR24259 standard; Protein; 689 AA.  
XX AC AAR24259;  
XX DT 09-NOV-1992 (first entry)  
XX Human amyloidin protease.  
XX DE  
XX KW Alzheimer's disease; beta amyloid precursor protein; APP; zinc;  
XX metalloprotease; hAP; protease inhibitor; ss.  
XX OS Homo sapiens.

Query Match 63.5%; Score 2331; DB 13; Length 689;  
Best Local Similarity 64.9%; Pred. No. 2.3e-192;  
Matches 426; Conservative 98; Mismatches 132; Indels 0; Gaps 0;

Qy 46 NVLRWDLSPQIKTTELLIVQTKQVDAVGMGLGTEETVYENCLOALADVVEKYIVERTM 105  
Db 22 nldrlwlsaqleertrelieqtkrvyqdvqtdqefedvsyestikaladvvetyvqrni 81

Qy 106 LDFPOHVSDEVRRASTEADKRLSRFDIEMSRGDIPIRVHLOETCDLGIKIRPEARRY 165  
Db 82 ldfpghvpskdirtasteadkklsefovemsredvgrivwlqekvqkdsirpeaary 141

Qy 166 LEKSTKMGKRNGLHLPQGVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKAEIGALPD 225  
Db 142 lerliklgrnglhlpretqenikrikkkllsldfknknlnedttflptqlglgipe 201

Qy 226 DFIDSLETKDDDKYITLKYPHYPMVKKCCIPETRRRMEAFNTRCKEENTIILOQLLP 285  
Db 202 dflnskmedgklkvtlkyphypllkchvpetrrkveafncrceenkaiklsvt 261

Qy 286 LRTKVAKLGYSTHADFLVLENTAKTSRVTAFDLDLSQKLKPLGEAREFRILNLKKEC 345  
Db 262 lraqksrllgfhthadyvlemmaktsgvatfidelakklpgeqeravilelakraec 321

Qy 346 KDRGFYDGKTNANDLYYWTQTTELKYSIDQEFLEKVEFPEVTEVTEGLNTYQELGLSF 405  
Db 322 egrglpfgdgrawdmrymqveetyrcvqdnllkeyfpvqvvtghllgylqellglaf 381

Qy 406 EOMTDAHVNKSVTLTYVKDKATCEVLGQFVLDLYPREGKYNHAACFGLPGCCLLPQGSR 465  
Db 382 Meegasawhedvrltyardasgevvgkfyldlypregyghaacfglqpgcrlrqdgsr 441

Qy 466 MMAVALVNFSPVAGRPSSLRHDEVTYFHERGHVHQICAQTDFAFSGTNNVETDFV 525  
Db 442 qiaiaamvanftkptadapsllqhdevetyhefghvnhqicsgaeafmgfsgthverdfv 501

Qy 526 EVPSQMLENWWVDLSRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLRQIVLSKVD 585  
Db 502 eapsgmlenwyweqepllrmarhyrtgsavprelleklesrqantgifsrlqvlakvd 561

Qy 586 QSLHTNTSLDAASEYAKYCEILGVAAATPGTNMPATFGHLAGGYDGYGYVLMSEVFSMD 645  
Db 562 qalhtqtadapaeayarlcqeilgvpatpgtntmpatfghlaggydaqygyvwsevymsmd 621

Qy 646 MFYSCFKKEGIMNPVGMKYNRLILKPGSLDGMDFLHFLKREPNAQFALMSRGL 701  
Db 622 mfhtrfkqevlnskvgmdyrscilrpggsedasamlrrfrlgrdpkqdaflilskgl 677

RESULT 4  
AAB93028  
ID AAB93028 standard; Protein; 381 AA.  
XX AAB93028;  
AC AAB93028;  
XX 26-JUN-2001 (first entry)  
DT Human protein sequence SEQ ID NO:11797.  
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
OS EPI074617-A2.  
PN 07-FEB-2001.  
PD 28-JUL-2000; 2000EP-0116126.  
XX 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX Claim 8; SEQ ID 11797; 2537pp + CD ROM; English.  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by



CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences: AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

Query Match	54.08;	Score 1981;	DB 22;	Length 381;
Best Local Similarity	95.29;	Pred. No. 1.5e-162;		
Matches 380;	Conservative 0;	Mismatches 1;	Indels 18;	Gaps 381
Qy	306	MNTAKSTRVAFADDDLSOKLPLGEAREFILLNKKKECKDRGFYDGKLNAWDLYYM	365	
Db	1	mntakstrvtafaddlsqsklpigearefilnlkkekckdrgefgydgklnawdlyyym	60	
Qy	366	TOTELKYSIDOEFLKEYFPTEVVTGELLNTYQELLGLSFGQMTDAHVMNKSVTLYTVKD	425	
Db	61	tqeelkysidqeflkeyfpievvtgellntyqellglisfeqmtdahvmksvtlytvkd	120	
Qy	426	KATEVILGQFVLDLYPRRGKYNHAACFGCLPLPGSRMVAALVNVNSQPVAGRPS	485	
Db	121	katgevilgqfylvldlyprgkynhaacfglpqclplpgsrmmvaalvvnvsqpvag-rps	180	
Qy	486	LLRHDEVNTYEPHEFGHVMHQICAOITDFARFSGTINVETDFEVPQSOMLENWMDVDSLRL	545	
Db	181	llrhde-----tdfarfsgtinvetdfvevpqsmlelenwvdvgsllrl	222	
Qy	546	SKHYKDGSPIDDLLEKIVASRLVNTGLLTIRQIVLSKVDQSLHTNTSLDAASEYAKYCS	605	
Db	223	skhykdgspiddllekiivasrlvntglltirlqvilskvddgslhtntsldaaseyakyCS	282	
Qy	606	BILGYAATPGTNMPATFGHLGAGGDGQYGYLWSEVFSMDMFYSCFKKEGTMPNPEVGMYK	665	
Db	283	eilgyaatpgtnmpatfghlaggydgqygyylwsevfsmdmfyescfkkegimpevgmky	342	
Qy	666	RNLILKPGGSLDGDMDLHNFLLKREPQNAFLMSRGLHAP	704	
Db	343	rnlilkpiggsldgmdlhnfllkrepnqkafimsglhap	381	

RESULT	5
AAB58575	
ID	AAB58575 standard; Protein; 678 AA.
XX	
AC	AAB58575;
XX	
DT	13-MAR-2001 (first entry)
XX	
DE	N. gonorrhoeae amino acid sequence g128.pep SEQ ID NO:53.
DE	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW	diagnosis; antigen; detection; infection; gene therapy; antibacterial.
XX	
OS	Neisseria gonorrhoeae.
XX	
PN	WO200066791-A1.
XX	
PD	09-NOV-2000.
XX	
PF	08-MAR-2000; 2000WO-US05928.
XX	
PR	30-APR-1999; 99US-0132068.
PR	08-OCT-1999; 99WO-US23573.
PR	28-FEB-2000; 2000GB-0004695.
XX	
PA	(CHIR ) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Pizza M, Hickey E, Peterson J, Tettelin H,
PI	Galeotti C, Mora M, Ratti G, Scarselli M,
PI	Venter JC, Scarlato V, Masignani V,
	Rappuoli R;

PI Frazer CM, Grandi G;  
XX WPI; 2000-647603/62.  
XX N-PSDB; AAF21570.  
XX  
XX Neisseria meningitidis B full length genome sequence and open reading  
PT frames are used to detect, treat and prevent Neisserial infections -  
XX  
XX Example 1; Page 93-94; 692pp; English.  
XX  
XX The present invention describes the full length genome of  
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607  
CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
CC sequence was too long to go in a record on its own it was split into 8  
CC sequences which overlap each other at the beginning and end of each  
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at  
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the  
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to  
CC AAF21606 represent PCR primers which are used in the exemplification of  
CC the present invention. The NMB genome and fragments from it have  
CC antibacterial activity, and can be used in vaccines and gene therapy.  
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
CC proteins can be used in compositions for treating or preventing infection  
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
CC bacteria. Computers, computer memory, computer storage medium or computer  
CC databases can be used in a search to identify open reading frames (ORFs)  
CC or coding sequences within the NMB genome. The DNA sequences provide  
CC further opportunities to find antigenic or immunogenic proteins which are  
CC more effective in vaccines  
XX  
XX Sequence 678 AA;  
XX



OS Neisseria gonorrhoeae.  
 PN WO200022430-A2.  
 PD 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 DR N-PSDB; AAA81285.  
 CC Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other Neisserial infections, for example, *N.gonorrhoea* -  
 XX Claim 14; Page 93; 1760pp; English.  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Meningococcus B*; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.  
 XX SQ Sequence 678 AA;

Query Match 19.1%; Score 702; DB 21; Length 678;  
 Best Local Similarity 28.7%; Pred. No. 8.9e-52;  
 Matches 194; Conservative 127; Mismatches 305; Indels 50; Gaps 18;

QY 56 QIKTRTEELIVOTKQYDVGMLGTEVT----YENCLQALADVEKVIYVERT----ML 106  
 DB 17 qiqtedikpavqt-aiaeargiaavkaqthtgawntverltg-----ltervgrigv 70  
 QY 107 DFPQHVSSDKVRAASTEADKRLSRFDIEMSRGDIIFERIVHLQETCDLGKIKPEARRYL 166  
 DB 71 shlnsvdtpelravynelpelvtvftteigqdielynrftktnkspefatlspaktkl 130  
 QY 167 EKSIRKNGKNGHLPEQVQNEIKSMKRMSELCDIPNKN-LNEDDTFLVF--SKAELGAL 223  
 DB 131 dhdlrfdvlsgealperqaelakiqtegaqisakfsqnlvdafatdgyfdadaaplagi 190  
 QY 224 PDD----FIDSLEKTDDDKYKITLKYPHYFPVPMKKCCIPETRRMEMAFNTRCKE----- 274

DB 191 pedalamfaaaqsegktykigqlqiphylavignrelreqiyrayvtraselnsdg 250  
 QY 275 --ENTIILOQLPLRTKVAKLIGYSTHADFLVEMNTAKTSRVTAFLDLSOKRLPLGEA 332  
 DB 251 kfdantaidrtlenalktakllgfknyaelslatkmdtpeqvlnflhlarrakpyaek 310  
 QY 333 EREFILNLKKCKDGRGFEYDGKIN--AWDLYYMTQTEELKYSIDOEFLKVFPIEVT 390  
 DB 311 d-----laevkafarehlgladqpqwdlsyagcklreakyafsetevkkyfpgkvl 362  
 QY 391 EGLLNTYOELLGLSFEQMTDAHWNKSVTLTYTKDKATGEVLGQFVLDLYPREGKYNHAA 450  
 DB 363 aglfaqikkllygigfaekt-vpwhkhkdvryfelqq-ngktiggvymdlyaregkrggaw 419  
 QY 451 CGLQLQCLLPDGRSMMAVAALVWVNSQVAGRPSSLRHDEVRTYFHERGHVHQICAQT 510  
 DB 420 mndykgrrrfadgtlqptayivcnfappvggkearlshdeltilfhetghghlhltdv 479  
 QY 511 DFAPTSCTN-VETDFEVEPVSQMLENWNWVDVSLRRLSKYKDGSPITADDLLEKLVASRLV 569  
 DB 480 delgvsingvewdaveipsgfmenfvweynvlagmsaheetgeplpkelfdkmlaaknf 539  
 QY 570 NTLTLTLRLQIVLSKVDQSLHTNTSLDAASEYAKYCYSEI---LGVAAATPGTNNMPA-TFGHL 625  
 DB 540 qrgmflvrqmfalfdmmyssedecrlknwqgvldsvrkevaviqppeynrnfansfghl 599  
 QY 625 -AGGYDCQYGYGLWSEVFSMDMFYSCFKKEGIMNPVGMKYRNLIILKPGSLDGMDLHN 684  
 DB 600 faggysagysyawaaveilstda-yaafesddv-aatgrfwgeillavdggsraaesfka 657  
 QY 685 FLKREPNOQAFILMSRG 700  
 DB 658 frgepsidallrqsg 673

RESULT 8  
 AAB25647  
 ID AAB25647 standard; Protein; 678 AA.  
 XX AAB25647;  
 AC AAB25647;  
 DT 04-DEC-2000 (first entry)  
 XX N. meningitidis amino acid sequence ml28-1.pap SEQ ID NO:1017.  
 DE Neisseria meningitidis; *Neisseria gonorrhoeae*; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW *Meningococcus B*; MenB.  
 XX *Neisseria meningitidis*.  
 OS *Neisseria meningitidis*.  
 XX WO200022430-A2.  
 PN 20-APR-2000.  
 PD 08-OCT-1999; 99WO-US23573.  
 PF 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizza M;  
 XX WPI; 2000-318079/27.  
 DR N-PSDB; AAA81287.  
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other Neisserial infections, for example, *N.gonorrhoea* -



Query Match	19.1%; Score 702; DB 21; Length 678;
Best Local Similarity	28.7%; Pred. No. 8.9e-52;
Matches	194; Conservative 127; Mismatches 305; Indels 50; Gaps 18;
QY	56 QIKTRTEELIVQKQYDVGMLGIEVT- -YENCLOALADVEVKYIVERT- -ML 106
Db	17 qidqtdikpavqt-alaeargaaiaaqaqthtgwantvertlg- -itervgrlwgvv 70
QY	107 DFPQHVSSDKEVRAASTADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPEARRYL 166
Db	71 shlnsvdtpelravynelpmeitvftfgeigdielynrftkiknspefatlspaqtkl 130
QY	167 EKSIMKGRNGLHLPQVQNEIKSMKRMSELICIDFNKN- -LNEDDTFLV- -SKAELGAL 223
Db	131 dhdldrfvlsgaeelpaqaeaklqtegaqlsakfsqnvldatdafgyfdaaplgi 190
QY	224 PDD- -FIDSLEKTDYKIKLTKYKPHYFPVMMKCCIPETRRRMENAFNTRCKE- - 274
Db	191 pedalamfaaaqsesktgykigilqlphylaviqyadnrelreqiyrayvtrasesldsg 250
QY	275 - -ENTIILOQLPLRTKVKALLGYSTHADVFLEMTAKTSRVTAFLDLSOKLPLGFA 332
Db	251 kfdntanidrtlanalqtakllgfknyaelstakmadtpeqvlnfhlldlarrakpyaek 310
QY	333 EREFILNLAKKCKDRGFYDQKIN- -AMDLYYYMTQTEELKYSIDQFLKYEYFPIEVVT 390
Db	311 d- - - - -laevkafarehlgldpqpwdlsayagekireakyaafsetevkyfpgkvl 362
QY	391 EGLLNTYQELLSFQMTDAHWKNSVTLTVKDKATGEVLGQPYLDLPYPRGKYNHAA 450
Db	363 agifaqikllylgfaekt-vpwwhkdvryfelqg- -ngktiggvymdlyaregrgaw 419
QY	451 CFGQLPGCLLPDGRMVAALVNFSPQVAGRPSLLRHDEVRTTFHFGVHWQICQAT 510
Db	420 mndykgrrrfadqtlqptlaylvcnfappvgvgearlsdelitfthgthghlltqv 479
QY	511 DPARFSGTN- -VEPDFVEVPSQMLENNWVDVSIIRLSKHYPKDGSPADLDLEKLVASRLV 569
Db	480 delgvsingvewadelpsfmenfvsynvlagmsaheetgeplkelfdkmlaaknf 539
QY	570 NTGLLTLRQILVSKVDQSLSHTNTSLDAASEYAKYGEI- - -LGVAATPQTNMPA- -TFGHL 625
Db	540 qrgmflvrqmfalfdmmyysesdecrlnkwwqvlsvrkevaviqpqpneynrfansfghi 599
QY	626 -AGGYDQYGYGLWSVFMDFYSCFKKEGIMNPEVGMKYNLLIKPGSLDGMMDMLHN 684
Db	600 fagysagysyawaevlstda-yaafesddv-aatgrfkwqeilavlggsrsaaesfka 657
QY	685 FLKREPNOKAFILMSRG 700
Db	658 frgrepsidallrqsg 673
RESULT	10
ID	AA74508
ID	AA74508 standard; Protein; 678 AA.
XX	AA74508;
AC	AA74508;
XX	21-MAR-2000 (first entry)
DT	Neisseria meningitidis ORF 128 protein sequence SEQ ID NO:490.
DE	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW	antibacterial; gene therapy.
XX	Neisseria meningitidis.
OS	Neisseria meningitidis.
XX	W09957280-A2.
PN	
XX	







Example 1; Page 96; 692pp; English.

The present invention describes the full length genome of Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613 represent fragments of the NMB genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins given in AAB38550 to AAB58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. Neisseria nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to Neisserial bacteria or as a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised to Neisserial bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines

Sequence 678 AA;

Query Match 18.9%; Score 694; DB 21; Length 678;  
Best Local Similarity 28.2%; Pred. No. 4.4e-51;  
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQITRTBELIVQT--KQYDAVGMIGTEVT-YENCLOALADVVKIYERT-----ML 106  
DB :||| : || : || : || : || : || : || : || : || : || : || : || :  
16 dqiktedkpalqtaiaeareglaaikaqtgthgwantvepltg----itervgirwgv 70

QY 107 DFPQHVSDEKVRAASTEADRLSRFEDIEMSMRGDIFERIVHLOETCDLGKIKPEARLYL 166  
DB :| : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
71 shlnsvtdtpeiraayneimpeitvffeigdqdlglynrfttknspedtlshaqktkl 130

QY 167 EKSTKMGRNGHLHPQOVNETSKMKRMSEICIDFNKN-LNEDDTFLVF--SKAELGA 223  
DB :|| : || : || : || : || : || : || : || : || : || : || : || :  
131 nhairdfvlsgaelppeqgaekalqttegaqlsakfsqnvdadafgyfdadaaplgi 190

QY 224 PDD----FIDLEKTDDDKYKITLYPHYFPVMKKCCIPETRRMEMAFNTCKE----- 274  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
191 pedalamfaaaagsegkgylgqlphylaviqyadrnkrlreqlyrayvtasealsddg 250

QY 275 --ENTIILOQLPLRTKVAKLLGYTHADFVLEMNTAKSTSVTAFLDDLSQLPLGEA 332  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
251 kfdntanidrtnenalgtaklglfknayaelslatkmdtpqvlfnflhdlarakpyaek 310

QY 333 EREFTLNKKKECKDRGPEYDGKINAWDLYYMTQTELKYSIDOEFLYEPPIEVVTEG 392  
DB :-|-laevkafarelsgl---adiqpwdlgvagelkreayafsetevkkyfpvgkylng 364

QY 393 LLNTYQELLGLGSFEQMTHAVWNKSVTLYTVKDKATGEVLGYFDLYPREGKYNHAACF 452  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
365 lfaqikklylgiftekt-vpvwhkavyfelqq--ngetiggvyndlyarekrgggawmn 421

QY 453 GLQPCLLPDGRMMAVAALVNFSQPVAGRPSLLRHDFVRYTFHEFGHVHQICAQTD 512  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
422 dykgrrrfsdgtqlptaylvcnfppvggkearlshdeilltfihetghhlhtqvde 481

QY 513 ARFCGTN-VETDFVBPVPSOMLENWVDLSRLRSKHXYKDGSPIADDLLEKLVASLVNT 571  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
482 lgvsingvewdavelpsqfmefnwynvlaqmsaheetgvplpkelfdkmlaanfr 541

QY 572 GILTLURQIVLSKVDOSLHTNTSLDAAASEYAKYCSET---LGVAAPTGTWNPA-TFGHL-A 626  
DB :|| : || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
542 cmfvromefaldmmyiseddegrikmgagvldsvirkvavvrpevirfansfhifa 601



Query Match	18.9%; Score 694; DB 21; Length 678;
Best Local Similarity	28.2%; Pred. No. 4.4e-51;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;	
QY 55	EQIKTRTEELIVQT--KQVYDVGMLGIEEVT-YENCLOALADVEVKYIVERT-----ML 106
Db 16	dqiktedikpalqataaeareqiaaikaqhtgtwantepltg-----itervgrivgvv 70
QY 107	DFPQHVSSEKVEAASTEADKRLSRDIEMSMRGDIPERIVHLOETCDLGKIKPEARRYL 166
Db 71	shlnsvtdtpelraaynelmpeltvftteigqdielynrftkiknspefdtlshaqtkl 130
QY 167	EKSIMGKRNGLHLPEQVQNEIKSMKRMSELICIDFNKN-LNEDDTFLVF--SKAELGAL 223
Db 131	nhdldrfvlsqaelpppeggaekalqtqegalsakfsqnlvdadafgyfddaaaplgi 190
QY 224	PDD----FIDSLEKTDQDDKYKITLKYPPVPMKCCIPETRRMEMAFNTRCKE----- 274
Db 191	pedalamfaaaasegktgkylqilqiphylavigyadnrklreqiyrayvtraseisddg 250
QY 275	--ENTILQOLLPLRTKVAKLGLGYSTHADVFLENMTAKSTSRTVAFLLDLSQKLPLGEA 332
Db 251	kfdntanidrtlenalqtakllgfknyaelslatkmadtpeqvlfnldlarrakpyaek 310
QY 333	EREFILNLKKCKDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKFFPIEVVTEG 392
Db 311	d----laevkafareslgl---adlqpdwldgyageklreaakyafsetevkyfpgvkvng 364
QY 393	LLNTYQELLGLSFEQMTDAHVWNKSVTLTVKDKATGCVLGGQFVLDLYPEGKYNHAACF 452
Db 365	lfaikllvgigttekt-vpvhkdvryfelq--ngetiggyvmdlyaregkrggawmn 421
QY 453	GLQPGCLLDGSRMAVAALVNFSPQVAGRPRLLRHDEVRTYFHEFGVHMQICAQDTDF 512
Db 422	dykgrfrfsdgtlqptaylvcnftppvggklearlshdeltilfhetghlhlhtqvde 481
QY 513	ARFSGTN-VETDFEVPQSOMLENWVDVSLRLSKHYKDGSPDIADLLEKLVASRLVNT 571
Db 482	lvsgslngvewdaveipsgmfewynvlagmsheetgvplpkelfdkmlaaknfr 541
QY 572	GLLTLROIVLKSDQSLHNTSLDAASEYAKYCSSEI---LGVAAATPGTNMPA-TFGHL-A 626
Db 542	gmflvrgmefaldmmylsdddegriknwqqvldsvrkevavvrpbeynrfansfghifa 601
QY 627	GGYDQGYGYLWSEVFSMDMFYSCPKKEGIMNPEVGMKYNRLILKPGGSLDGMMLHNF 686
Db 602	gysagysyawaevlsada-yaafesddv-aatgkrfwgeilavggssraaesfkafr 659
QY 687	KREPNOKAFILMSRG 700
Db 660	grepsidallrhsg 673
RESULT 15	
AA25646	
ID	AA25646 standard; Protein; 678 AA.
XX	
AC	AA25646;
XX	
DT	
XX	
XX	04-DEC-2000 (first entry)
DE	N. meningitidis amino acid sequence al28.pep SEQ ID NO:1015.
XX	
KW	Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW	antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW	Meningococcus B; MenB.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO200022430-A2.
XX	

20-APR-2000.

08-OCT-1999; 99WO-US23573.

09-OCT-1998; 98US-0103794.

30-APR-1999; 99US-0132068.

(CHIR ) CHIRON CORP.

Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC; Masignani V, Galicotti C, Mora M, Ratti G, Scarselli M, Scarlato V; Rappuoli R, Pizza M;

WPI: 2000-318079/27.

N-PSDB; AA81286.

Isolated nucleotide sequences of *Neisseria meningitidis* which can be used in the diagnosis and treatment of *N. meningitidis* infection and other *Neisseria* infections, for example, *N. gonorrhoea* -

Claim 14; Page 95; 1760pp; English.

The present invention describes methods of obtaining immunogenic proteins from *Neisseria* genomic sequences. AA81453 to AA82414 represent specifically claimed *Neisseria meningitidis* genomic DNA sequences; AA81260 to AA81303 and AA825620 to AA825663 represent *Neisseria* DNA sequences and their corresponding proteins; AA81254 to AA81259 and AA81304 to AA81321 represent PCR primers used in the isolation of *Neisseria meningitidis* DNA sequences; and AA81322 to AA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all used in the exemplification of the present invention. The nucleic acid sequences, protein sequences, and antibodies against them, can be used in the manufacture of a composition. The composition can be used as a medicament (or in the manufacture of a medicament) for treating, preventing or diagnosing infection due to *Neisseria* bacteria. For example, some of the identified proteins could be components of vaccines against *Meningococcus B*; against all serotypes; and/or against all pathogenic *Neisseriae*. Identification of sequences from the bacterium will also facilitate production of biological probes, particularly organism-specific probes. Attempts to make efficacious *Meningococcus B* vaccines have failed mainly due to antigen tolerance. Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and which are not antigenically variable or at least more conserved than other more variable regions.

Sequence 678 AA;

Query Match 18.9%; Score 694; DB 21; Length 678;

Best Local Similarity 28.2%; Pred. No. 4.4e-51;

Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQIKTRTEELIVQT--KQVYDVGMLGIEEVT-YENCLOALADVEVKYIVERT-----ML 106

Db 16 dqiktedikpalqataaeareqiaaikaqhtgtwantepltg-----itervgrivgvv 70

QY 107 DFPQHVSSEKVEAASTEADKRLSRDIEMSMRGDIPERIVHLOETCDLGKIKPEARRYL 166

Db 71 shlnsvtdtpelraaynelmpeltvftteigqdielynrftkiknspefdtlshaqtkl 130

QY 167 EKSIMGKRNGLHLPEQVQNEIKSMKRMSELICIDFNKN-LNEDDTFLVF--SKAELGAL 223

Db 131 nhdldrfvlsqaelpppeggaekalqtqegalsakfsqnlvdadafgyfddaaaplgi 190

QY 224 PDD----FIDSLEKTDQDDKYKITLKYPPVPMKCCIPETRRMEMAFNTRCKE----- 274

Db 191 pedalamfaaaasegktgkylqilqiphylavigyadnrklreqiyrayvtraseisddg 250

QY 275 --ENTILQOLLPLRTKVAKLGLGYSTHADVFLENMTAKSTSRTVAFLLDLSQKLPLGEA 332

Db 251 kfdntanidrtlenalqtakllgfknyaelslatkmadtpeqvlfnldlarrakpyaek 310

QY 333 EREFILNLKKCKDRGFEDGKINAWDLYYMTQTEELKYSIDQEFLEKFFPIEVVTEG 392

Db 311 d----laevkafareslgl---adlqpdwldgyageklreaakyafsetevkyfpgvkvng 364

QY 393 LLNTYQELLGLSFEQMTDAHVWNKSVTLTVKDKATGCVLGGQFVLDLYPEGKYNHAACF 452

Db 365 lfaikllvgigttekt-vpvhkdvryfelq--ngetiggyvmdlyaregkrggawmn 421

QY 453 GLQPGCLLDGSRMAVAALVNFSPQVAGRPRLLRHDEVRTYFHEFGVHMQICAQDTDF 512

Db 422 dykgrfrfsdgtlqptaylvcnftppvggklearlshdeltilfhetghlhlhtqvde 481

QY 513 ARFSGTN-VETDFEVPQSOMLENWVDVSLRLSKHYKDGSPDIADLLEKLVASRLVNT 571

Db 482 lvsgslngvewdaveipsgmfewynvlagmsheetgvplpkelfdkmlaaknfr 541

QY 572 GLLTLROIVLKSDQSLHNTSLDAASEYAKYCSSEI---LGVAAATPGTNMPA-TFGHL-A 626

Db 542 gmflvrgmefaldmmylsdddegriknwqqvldsvrkevavvrpbeynrfansfghifa 601

QY 627 GGYDQGYGYLWSEVFSMDMFYSCPKKEGIMNPEVGMKYNRLILKPGGSLDGMMLHNF 686

Db 602 gysagysyawaevlsada-yaafesddv-aatgkrfwgeilavggssraaesfkafr 659

QY 687 KREPNOKAFILMSRG 700

Db 660 grepsidallrhsg 673

RESULT 15

AA25646

ID AA25646 standard; Protein; 678 AA.

XX

AC AA25646;

XX

DT

XX

XX 04-DEC-2000 (first entry)

DE N. meningitidis amino acid sequence al28.pep SEQ ID NO:1015.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KW Meningococcus B; MenB.

XX

OS Neisseria meningitidis.

XX

PN WO200022430-A2.

XX

Db 251 kfdntandirtlenalqtakllgfknyaelslatkmdtpeqvlnflhdlarrakpyaek 310  
QY 333 BREFILNLKKCKEKGFEYDCKINAWDLYYMYTQTEELKYSIDOEFLKEYPPIEVWTEG 392  
Db 311 d---laevkafareslgl---adlcpwdlgyageklreakyafsecevkkyfpvgkvlnq 364  
QY 393 LNTTYQELLGLSFEQMTDAHVWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAACF 452  
Db 365 lfaqlkllygigftek-vpwhkdvryfelq--ngetigvymdlyaregkrggawmn 421  
QY 453 GLQPGCLLPDGSRMMAVALVNFQPVAGRPSSLRHDEVRTYFHEFGHVMHQICAQTDf 512  
Db 422 dykgrrrfsdgtlqiptaylvncfppvggkearlshdeilltfhetghlhhlltqvde 481  
QY 513 ARFSGTN-VETDFVEVPQMLENWWVDLSRLSKHYKDGSPiADDDLEKLVASRLVNT 571  
Db 482 lqvsqngvewdaveipsgfmenfweynvlaqmsaheetgvplpkelfdkmlaaknqfr 541  
QY 572 GLTLRLQIVLSKVDQSLHTNTSLDAASEYAKYCSEI---LGVAATPGTNMPA-TFGHL-A 626  
Db 542 gmflvrqmfalfdmniyseddegrlknwqqlvdsrkevavvrppeynrfansfghifa 601  
QY 627 GGYDQGYGYLWSEVFSMDMFYSCFKKGIMNPEVGMKYRNLIILKPGSLDGMMLHNFL 686  
Db 602 ggyysagyysawaevlsada-yaafeesddv-aatgkrfwqellavggssrsaaesfkafr 659  
QY 687 KREPNOKAFLMSRG 700  
Db 660 grepsidallrhsg 673

Search completed: January 14, 2002, 18:55:59  
Job time: 1043 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:00:21 ; Search time 91.83 Seconds  
(without alignments)  
5216.171 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatcccggtgcctttt.....gagcgctgcgtcgtgga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	%	Length	ID	Description
1	1590.8	75.2	2148	5	PCT-US92-00731-12	Sequence 12, Appl
2	790.8	37.4	2070	1	US-07-766-351-6	Sequence 6, Appl
3	790.8	37.4	2070	1	US-08-059-032-6	Sequence 6, Appl
4	790.8	37.4	2070	5	PCT-US91-07290-6	Sequence 6, Appl
5	86.8	4.1	723	4	US-08-998-416-839	Sequence 839, Appl
6	61	2.9	7218	1	US-08-232-463-14	Sequence 14, Appl
7	42	2.0	75	1	US-07-766-351-13	Sequence 13, Appl
8	42	2.0	75	1	US-08-059-032-13	Sequence 13, Appl
9	42	2.0	75	5	PCT-US91-07290-13	Sequence 13, Appl
10	41.4	2.0	164	1	US-07-766-351-16	Sequence 16, Appl
11	41.4	2.0	164	1	US-08-059-032-16	Sequence 16, Appl
12	41.4	2.0	164	5	PCT-US91-07290-16	Sequence 16, Appl
13	36.2	1.7	1821	4	US-09-056-105-23	Sequence 23, Appl
14	36.2	1.7	1896	1	US-08-253-503-1	Sequence 1, Appl
15	36.2	1.7	1896	1	US-08-796-883-1	Sequence 1, Appl
16	36.2	1.7	1896	2	US-08-611-273B-1	Sequence 1, Appl
17	36.2	1.7	1896	2	US-08-531-864-1	Sequence 1, Appl
18	36.2	1.7	1896	2	US-08-373-636C-1	Sequence 1, Appl
19	36.2	1.7	1896	3	US-08-602-506A-1	Sequence 1, Appl
20	36.2	1.7	1896	4	US-09-266-294-1	Sequence 1, Appl
21	36.2	1.7	1896	4	US-09-179-281-1	Sequence 1, Appl
22	36.2	1.7	1896	4	US-09-056-105-25	Sequence 25, Appl
23	35.4	1.7	3095	6	5231168-1	Patent No. 5231168
24	35.2	1.7	5741	1	US-07-706-699-4	Sequence 4, Appl
25	35.2	1.7	5741	1	US-07-998-931-4	Sequence 4, Appl
26	34.4	1.6	651	2	US-08-929-418-1	Sequence 1, Appl
27	34.2	1.6	5049	1	US-08-336-345-1	Sequence 1, Appl

Sequence 2, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 23, Appl  
Sequence 1, Appl  
Sequence 29, Appl  
Sequence 1, Appl  
Sequence 23, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 5, Appl  
Sequence 3, Appl  
Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
PCT-US92-00731-12  
; Sequence 12, Application PC/TUS9200731  
; GENERAL INFORMATION:  
; APPLICANT: Kawabata, Shunichiro  
; APPLICANT: Davie, Earl W.  
; TITLE OF INVENTION: MICROSOMAL ENDOPEPTIDASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00731  
; FILING DATE: 19920128  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/646,997  
; FILING DATE: 28-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 990008.415PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)622-4900  
; TELEFAX: (206)682-6031  
; TELEX: 3723836  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2148 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FRAGMENT TYPE: N-terminal  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 11..2075  
; PCT-US92-00731-12

Query Match 75.2%; Score 1590.8; DB 5; Length 2148;

[illegible]





QY 1028 aggaatgcaaaacagaggggttttgaatatgatgggaaatcaatgcctggggtatatt 1087  
Db 956 CGGAGTGCAGGCGCGGCGCTGCGCTTCGACGGCGCATCGCTGCCGTGGACATCGGCT 1015  
QY 1088 actacatgactcagacagaggaactcaagtatctcatagaccagaagttcctcaaggaa 1147  
Db 1016 ACTATGTAACCAAGTGGAGGAGACGCGCTACTGCGTGACCAACACCTGCTCAAGGAGT 1075  
QY 1148 acttcccaattgagtggtcactgaaggttctgtgaacacacctaccagaggttctgggac 1207  
Db 1076 ACTTCCCGTGAGTGGTTCACGACGGCTGCTGGGCTATCTACAGGAGCTCCTGGGGC 1135  
QY 1208 ttctattgaaacaatgacagatgctctgttttggaaagaagtgttacactttatctg 1267  
Db 1136 TGGCCTTCCACCACGAGGAGGCGCGAGTGGCTGCGATGAGGACCTGCGGCTCTACACG 1195  
QY 1268 tgaagataaactcagagagaagtatttggacagttctatttgaacctctatccaaggg 1327  
Db 1196 CGAGGACCGCGCTCGGGGAGGTGGTGGCAAGTTCTACCTGGACCTGTACCCGCGGG 1255  
QY 1328 aaggaaatacaatcatcgcgctctgtctcagctgctgctcttgcctgatg 1387  
Db 1256 AGGAAAGTACGGGACGCGGCTGCTTGGCTCTGACCGCGCTGCTGGCGGAGGATG 1315  
QY 1388 gaagccggatgagtgagtggtgctgctgctgtgtgaacttctcacagccagtggcaggtc 1447  
Db 1316 GGAGCGCGCAGATCGCATCGCGGCGCATGTGGCCAACTTCAACCAAGCCACAGCGGAG 1375  
QY 1448 gtccctctctctgagacagacagaggtgaggaacttctatagtttggctcagtg 1507  
Db 1376 CCGCTCTGCTGTGAGCATGACGAGGTGGAGACCTACTTCCATGAGTTTGGCCACGTA 1435  
QY 1508 tgcatcagattgtgcacagactgatttgcacgatttagcggaaacaagtgggaaactg 1567  
Db 1436 TGCACAGCTCTGTCCCGGCGGAGTTCGCCATGTTACGCGGGACCCACGTTGGAGCGGG 1495  
QY 1568 actttgtagagtgccatgcgcaaatgcttgaataattgggtggtggagctcgatccctcc 1627  
Db 1496 ACTTTGTGGAGCGCGCTGCGAGATGCTGGAGAACTGGGTGGGAGGAGCGGCTGC 1555  
QY 1628 gaagattgtcaaaacattataaagatgaagccctatttgagacgatctgtgaaaaac 1687  
Db 1556 TCGGATGTGCGGGCACTACCCACAGGACGCGCGCTGCGCGGAGGCTCCTGGAGAAAG 1615  
QY 1688 ttgttctcttagctgttcaacagacttcttgaacctgcgacagattgttttgagca 1747  
Db 1616 TCATTGAGTCCCGGAGGCGCAACACAGGCGCTTTCAGCGCTGCGCCAGATCGCTCGCCA 1675  
QY 1748 aagttgatcagttctctcatcaacaacacatcgctggtgctgcaagtgaaatgcaaat 1807  
Db 1676 AGGTGGACCAAGCGCTTGACACGCGAGCGGACGCGACCCCGGAGGATGCGCGGC 1735  
QY 1808 actgtcagaaaattagagttgagctacttccaggcacaataattgcagctactttg 1867  
Db 1736 TCTGCCAGGAGATCTCGGGTCCCGGCCACGCCAGGAACAACATGCTGCAACCTCG 1795  
QY 1868 gacatttggcaggggatacagatgcccataattatgatatcttttggagtgaaagtattt 1927  
Db 1796 GCATCTGCGAGGTGGCTACGACGCGCCAGTACTACGGGTACTCGGTGAGCGGAGGTGATT 1855  
QY 1928 ccattgatattgtttacagctgtttttaaagaagaggataatgaatccaaggttgaa 1987  
Db 1856 CCATGGACATGTTCCACACGCGCTTCACGAGGAGGGTGTCTCTCAACAGCAAGGTTGCA 1915  
QY 1988 tgaatacagaaacctaacttgaacctgggggagctctctgacggcagtggaatgctcc 2047  
Db 1916 TGGATTACAGAAAGCTGATCTCTGACCGCGGGTTCGAGGATGCCAGCGCCATGCTGA 1975  
QY 2048 acaatttcttgaacgtagacaaacaaagcgttctaatgagtagagcctgcatg 2107  
Db 1976 GCGGCTTCTTGGGCGGTGACCCCAAGCAGGAGCGCTTCTCTGAGCAAGGGGCTGCAAG 2035

RESULT 4  
PCT-US91-07290-6  
; Sequence 6, Application PC/TUS9107290  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07290  
; FILING DATE: 19911004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Felix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2070 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US91-07290-6

Query Match 37.4%; Score 790.8; DB 5; Length 2070;  
Best Local Similarity 62.4%; Pred. No. 7.3e-217;  
Matches 1236; Conservative 1; Mismatches 743; Indels 0; Gaps 0;  
QY 128 ctggcagaaagtgtttaagatggggtatctttccaccagagcaaatataaagaactgagg 187  
Db 56 CTGTGTTAAACACCTCGGTGGGACCTGAGTGCCACGATAGAGGAGCGCACGAGG 115  
QY 188 agctcattgtgcagaccacaaaggtgtgacgatgctgttggaaatgctcggattgaggaag 247  
Db 116 AGCTCATCGACGACACCAAGCGGTGTATGACCAAGTTGGCACCCAGGAGTTGAGGAG 175  
QY 248 taacttacgaaactgtctgcaggcactggcagatgtagaagtaaatatagtggaag 307  
Db 176 TGTCTACGACGACGCTCAAGGCGCTGGCGGATGTGGAGTCACTACACAGTTTCA 235  
QY 308 ggaccatgtagactttccccagcatgtatctctgacaaagaagtacgagcagcaagta 367  
Db 236 GGAATATCTTGTACTTCCCGCAGCATGTTTCCCGCTCCAAAGGACATCCGACACGCA 295  
QY 368 cagaagcagacaaagaactttctctgttttggatattgagatgagcagagagatatat 427  
Db 256 CAGAGGCGGACAAAGAGCTCTGTGAGTTTCGAGCTGGAGATGAGGAGGAGCGGTGT 355  
QY 428 ttgagagaattgtttcatttacagggaacactgtgatctgggggaagataaaacctgagcc 487

Db 356 ACCAGAGATCGTGTGGCTCCAGGAGAAAGTTCAAGAGACTCACTAGGCGCCGAGGCTG 415  
Qy 488 gacgatacttgagaaagtcattaaattgggagaaagaaatgggctccatcttccctgaac 547  
Db 416 CGCGTACTTGAGGGGCTAATCAAGCTGGCGGAGAAATGGGCTTCACCTCCCCAGAG 475  
Qy 548 aaglacagaatgaatacaataatgaagaaagaaatgagtgagctatgtattgatttta 607  
Db 476 AGACTCAGGAAACATCAACGCATCAGAGAGAACCTGAGCCTTCTGTGCATGCATCA 535  
Qy 608 acaaaaaacctcaatgagatgataaccttcccttctgtattttccaaggctgaacttggctgc 667  
Db 536 ACAAGAACTGAAGAGAGACAGCAGCTTCTTCCCTTTCACGCTCCAGGAGCTAGGAGGC 595  
Qy 668 ttctgtgatttcatgacagttttaaaaaagacagatgacagaagtataaaattacct 727  
Db 596 TCCCGAGGACTTCTGNACTTCCTTGAAGAGATGGAGAGCGGAAGTTGAAGGTCAACC 655  
Qy 728 taataatccacactatttccctgctcatgaagaaatgttgtatccctgaaaccagaagaa 787  
Db 656 TCAAGTACCCCATTAATTCCCTCCCTCTGAAGAAATGCCACGTGCTGAGACAGGAGGA 715  
Qy 788 99atggaataggcttttaatacaagggtgcaagagggaaaaaacacataattttgcagcagc 847  
Db 716 AAGTGGAGGAGGCTTCAACTGCCGGTGCAGAGGAGAGAACTGGCGCTATCCTCAAGGAGC 775  
Qy 848 taetcccaactgcgaacaaagtggccaaactactcgtgttatagcacacatgtaacttcg 907  
Db 776 TGGTGACCTCGCGGCCCCAGAAAGTCCCGCTGCTGGGGTTCCACAGCAGCCGACACTATG 835  
Qy 908 tccttgaatgaacactgcaagagacaaagccgcgtaacagcyctttttagatgatttaa 967  
Db 836 TCCTGGAGATGAACATGSCCAAGACCAAGACCGACCGTGGCCACCTCTCTAGATGAGCTGG 895  
Qy 968 gccagaaagttaaacaccttgggtggaagcagacagagagtttatttgaatttgaagaaaa 1027  
Db 896 CGCAAGACTGAAGCCCTCGGGGGAGCAGGAGCGTGGCGTGTCTGAGAGTGGAGCGGTG 955  
Qy 1028 aggaatgcaagacaggggttttgaatatgatgggaaatcaatgcctggtatatt 1087  
Db 956 CGGAGTGGACGCGCGGGGCTCGCCTTCGACGCGCGCATCCGTGCTGGGACATGCGCT 1015  
Qy 1088 actacatgactcagacagagaaactcaagtatttccatagaccagagttctcgaagaa 1147  
Db 1016 ACTACATGAACCAAGCTGGAGGAGACGCGCTACTGCTGGACCAACCTGCTCAAGGAGT 1075  
Qy 1148 acttcccaatgaggtggtcactgaagcttgcgtgaacacactacacagagttgtggac 1207  
Db 1076 ACTTCCCGCTGACAGTGGTGCACGACCGGCTGCTGGGCATCTACACGAGAGCTCCTGGGGC 1135  
Qy 1208 ttcatattgaacaaatgacagatgctcatgtttggaacaaagtggttacacttttactg 1267  
Db 1136 TGGCTTCCACCAAGAGGAGGGCCAGTGGCTGGCATGAGAGCTGGCGGCTCTACACCG 1195  
Qy 1268 tgaaggataaagctacagggagaagatttgggacaggtttctatttggactctatccaagg 1327  
Db 1196 CGAGGAGCGCGCTCGGGGAGGTGGTGGCAAGTCTTACTTGACCTGATACCCGCGG 1255  
Qy 1328 aaggaaatacaataatgcggcctgcttgcgttccagcctgctgccttgcctgatg 1387  
Db 1256 AAGGAAATAGCGGACGCGGCTGCTTGGCTTGCAGCCGCGCTGCTGGCGAGGATG 1315  
Qy 1388 gaagccgagatgagcagtgagctccctctggttggaacttctcagccagtgagcagtc 1447  
Db 1316 GGAGCGCGCATCGGCATCGCGCCATGGTGGCCACTTCCCAAGCCCAACAGCCGAGC 1375  
Qy 1448 gtccctctctctgagacagacagagtgagagcttactttcatgagtttggtcacatga 1507  
Db 1376 CGCCTCGCTGCTCAGCATGACAGAGTGGAGACCTACTTCCATGAGTTGGCCACATGA 1435  
Qy 1508 tgcatacgatttgcacagactgattttgcacagatttagcggaacaaatgtggaaactg 1567  
Db 1436 TGCACAGCTCTGCTCCAGCGGAGTTCCGCAATGTTCAAGCGGGACCCAGCTGGAGCGGG 1495

Qy 1568 actttagaggtgccatcgcaaatgcttgaataatgggtgtggagctgcattccctcc 1627  
Db 1496 ACTTTGTGGAGGCCCGCTCGCGATGCTGGAGAACTGGGTGTGGAGCAGAGCGGCTGC 1555  
Qy 1628 gaagattgtcaaaacattataaagatgaaagccctatttgcagacgactgtcttgaanaac 1687  
Db 1556 TCGGATGTCGCGGCACTTACCGCACAGCAGCGCGTCCCGGAGCTCCTTGAGAAAGC 1615  
Qy 1688 ttgttcttagtctggttcaaacacagcttcttgcacctgcgcagattgttttggagca 1747  
Db 1616 TCATTGAGTCCCGCAGGCCAACACAGGCCCTTTCAGCCTGCGCCAGATCCTCTCGCCA 1675  
Qy 1748 aagttgatcagctcttctcatcaacacacatcgtggtgctgcaagtgaatataatgcaaat 1807  
Db 1676 AGTGGACCAAGCCCTGACACAGCAGCGGACGAGACCCCGCCGAGGAGTATGCGCGGC 1735  
Qy 1808 actgctcagaataattaggagttgcagctactcctcaggcacaataatgccagctaccttg 1867  
Db 1736 TCTGCAGGAGATCCTCGGGTCCCGGCCACGCGCAGGAACCAACATGCTGCAACCTTCG 1795  
Qy 1868 gacatttggcaggggatacagatggccaataattatgatatcttcttggagtgaaatttt 1927  
Db 1796 GCCATCTGGCAGGTGGCTACGACGCCAGTACTACGGGTACTCTGGAGCGAGGTGTAAT 1855  
Qy 1928 ccattgatatatttgcagctgttttaaaaaagagggataatgaatccagaggttgaa 1987  
Db 1856 CCATGGACATGTTCCACACGCGCTTCAAGCAGAGGGGTGCTCTGAACAGCAAGGTGGCA 1915  
Qy 1988 tgaatacagaaacctaatcctgaacctggggatctcttggcggcagtgacatgctcc 2047  
Db 1916 TGGATTACAGAAGTGCATCTGAGACCGCGGTTCGAGGATGCCAGCGCATGCTGA 1975  
Qy 2048 acaatttctgaaacgtgagcccaacaaagaggttcttcttaataatgtagtaggctgcatg 2107  
Db 1976 GCGCTTCTTGGCGGTGACCCCAAGCAGGACGCTTCTCTCTGAGCAAGGGGCTGCAGG 2035

## RESULT 5

US-08-998-416-839

: Sequence 839, Application US/08998416

: Patent No. 6239264

: GENERAL INFORMATION:

: APPLICANT: Philippsen, Peter

: APPLICANT: Pohlmann, Rainer

: APPLICANT: Steiner, Sabine

: APPLICANT: Mohr, Christine

: APPLICANT: Wendland, Jurgen

: APPLICANT: Knechtle, Philipp

: APPLICANT: Reibschung, Corinne

: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPI

: TITLE OF INVENTION: AND USES THEREOF

: NUMBER OF SEQUENCES: 1152

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: No. 6239264artis Corporation

: STREET: 3054 Cornwallis Road

: CITY: Research Triangle Park

: STATE: No. 6239264th Carolina

: COUNTRY: USA

: ZIP: 27709

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/998,416

: FILING DATE: 24-DEC-1997

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: CH 0016/97

: FILING DATE: 31-DEC-1996

: ATTORNEY/AGENT INFORMATION:





Db 1085 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCAGCAA 1039

RESULT 7 |  
US-07-766-351-13  
? Sequence 13, Application US/07766351  
? Patent No. 5292652  
? GENERAL INFORMATION:  
? APPLICANT: Sinha, Sukanto  
? APPLICANT: Seubert, Peter A.  
? APPLICANT: Dovey, Harry F.  
? APPLICANT: McConlogue, Lisa C.  
? APPLICANT: Little, Sheila P.  
? APPLICANT: Johnstone, Edward M.  
? TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
? NUMBER OF SEQUENCES: 18  
? CORRESPONDENCE ADDRESS:  
? ADDRESSEE: Athena Neurosciences, Inc.  
? STREET: 800F Gateway Blvd.  
? CITY: South San Francisco  
? STATE: California  
? COUNTRY: USA  
? ZIP: 94080

```
Query Match: 2.0%; Score 42; DB 1; Length 75;
Best Local Similarity 73.0%; Pred. No. 0.0055;
Matches 54: Conservative 0; Mismatches 20; Indels 0; Gaps 0;
```

Qy	1141	aaggaatacttcccaattgagggtggtcactgaagccttgctgaacacctaccaggagttg	1200
Dp	1	AAGGAGTACTTCCTGTGGAGGTGGTCAGCACGGCTGCTGGGCATCTACAGGAGGTC	60

Qy	1201	ttgggactttcatt	1214
Db	61	CTGGGCCCTGGCCCTT	74

RESULT 8  
US-08-059-032-13  
; Sequence 13, Application US/08059032  
; Patent No. 5424305  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.

APPLICANT: McConlogue, Lisa C.  
 APPLICANT: Little, Sheila P.  
 APPLICANT: Johnstone, Edward M.  
 TITLE OF INVENTION: Anyloidin Protease and Uses Thereof  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Townsend and Townsend Kourie and Crew  
 STREET: One Market Plaza, Steuart Street Tower,  
 STREET: Suite 2000  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94105  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/059,032  
 FILING DATE: 19930507  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Smith, William E.  
 REGISTRATION NUMBER: 30,223  
 REFERENCE/DOCKET NUMBER: 15270-10  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 326-2400  
 TELEFAX: (415) 326-2422  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 75 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 HYPOTHETICAL: YES  
 ANTI-SENSE: NO  
 US-08-059-032-13

Query Match 2.0%; Score 42; DB 1; Length 75;  
Best Local Similarity 73.0%;  
Pred. No. 0.0055;  
Matches 54: Conservative 0; Mismatches 20; Indels

Qy	1141	aaggataacttcccaattgagggtggtcactgaagccttgc	gaacacctaccaggaggttg	1200
Dh	1	aaggcactacttccctctgacagctgctcacacacggcctgcttggc	atctaccaggaggtc	60

QY	1201	ttgggactttcatt	1214
Dh	61	CTGGGCTGGCCTT	74

RESULT 9  
PCT-US91-07290-13  
Sequence 13, Application PC/TUS9107290  
GENERAL INFORMATION:  
APPLICANT: Sinha, Sukanto  
APPLICANT: Seubert, Peter A.  
APPLICANT: Dovey, Harry F.  
APPLICANT: McConlogue, Lisa C.  
APPLICANT: Little, Sheila P.  
APPLICANT: Johnstone, Edward M.  
TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Athena Neurosciences, Inc.  
STREET: 800F Gateway Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

```

;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: PCT/US91/07290
;
; FILING DATE: 19911104
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Murphy, Lisabeth Rex
;
; REGISTRATION NUMBER: 31547
;
; REFERENCE/DOCKET NUMBER: 17796-002.
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (415) 877-0900
;
; TELEFAX: (415) 877-8370
;
; INFORMATION FOR SEQ ID NO: 13:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 75 base pairs
;
; TYPE: NUCLEIC ACID
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA
;
; HYPOTHETICAL: YES
;
; ANTI-SENSE: NO
;
; PCT-US91-07290-13

```

RESULT 10  
US-07-766-351-16  
; Sequence 16, Application US/07766351  
; Patent No. 5292652  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/766,351  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547

```

; REFERENCE/DOCKET NUMBER: 17796-002
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 877-0900
; TELEFAX: (415) 877-8370
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 164 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-07-766-351-16

```

```

RESULT 11
US-08-059-032-16
  Sequence 16, Application US/08059032
  Patent No. 5424205
  GENERAL INFORMATION:
  APPLICANT: Sinha, Sukanto
  APPLICANT: Seubert, Peter A.
  APPLICANT: Dovey, Harry F.
  APPLICANT: McConlogue, Lisa C.
  APPLICANT: Little, Sheila P.
  APPLICANT: Johnstone, Edward M.
  TITLE OF INVENTION: Amyloidin Protease and Uses Thereof
  NUMBER OF SEQUENCES: 18
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Townsend and Townsend Khourile and Crew
  STREET: One Market Plaza, Steuart Street Tower,
  STREET: Suite 2000
  CITY: San Francisco
  STATE: California
  COUNTRY: USA
  ZIP: 94105
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/059,032
  FILING DATE: 19930507
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
  NAME: Smith, William E.
  REGISTRATION NUMBER: 30,223
  REFERENCE/DOCKET NUMBER: 15270-10
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 326-2400
  TELEFAX: (415) 326-2422
  INFORMATION FOR SEQ ID NO: 16:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 164 base pairs

```

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-059-032-16

Query Match 2.0%; Score 41.4; DB 1; Length 164;  
Best Local Similarity 56.1%; Pred. No. 0.012;  
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 775 gaacacagaagatggaaatgcttttaatacaaggtgcaaaagaggaacacacata 834  
Db 26 GGAAGTGGAGGAGGCTTCAACTGCCGGTCAAGCGGTGCAAGGAGGAGAACTGGCT 85  
Qy 835 atttgcagcagctactcccactgcgaacaaagtgccaaactactcgtttatagaca 894  
Db 86 ATCTCAAGGAGCTGGTGACGCTCGGGGCCAGAAAGTCCCGCTGCTGGGGTTCACACG 145  
Qy 895 catgctgacttcgctctg 913  
Db 146 CACGCCGACTATGCTCTGG 164

RESULT 12  
PCT-US91-07290-16  
; Sequence 16, Application PC/TUS9107290  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/07290  
; FILING DATE: 19911004  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 164 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US91-07290-16

Query Match 2.0%; Score 41.4; DB 5; Length 164;  
Best Local Similarity 56.1%; Pred. No. 0.012;  
Matches 78; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
Qy 775 gaacacagaagatggaaatgcttttaatacaaggtgcaaaagaggaacacacata 834  
Db 26 GGAAGTGGAGGAGGCTTCAACTGCCGGTCAAGCGGTGCAAGGAGGAGAACTGGCT 85  
Qy 835 atttgcagcagctactcccactgcgaacaaagtgccaaactactcgtttatagaca 894  
Db 86 ATCTCAAGGAGCTGGTGACGCTCGGGGCCAGAAAGTCCCGCTGCTGGGGTTCACACG 145  
Qy 895 catgctgacttcgctctg 913  
Db 146 CACGCCGACTATGCTCTGG 164  
RESULT 13  
US-09-056-105-23/c  
; Sequence 23, Application US/09056105  
; Patent No. 6287569  
; GENERAL INFORMATION:  
; APPLICANT: KIPPS, THOMAS J.  
; APPLICANT: WU, YUNQI  
; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR  
; TITLE OF INVENTION: PROCESSING  
; FILE REFERENCE: 233/221  
; CURRENT APPLICATION NUMBER: US/09/056.105  
; CURRENT FILING DATE: 1998-04-06  
; EARLIER APPLICATION NUMBER: 60/043,467  
; EARLIER FILING DATE: 1997-04-10  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 23  
; LENGTH: 1821  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-056-105-23

Query Match 1.7%; Score 36.2; DB 4; Length 1821;  
Best Local Similarity 55.0%; Pred. No. 1.3;  
Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
Qy 486 cagacgatacttgaaagtcaattaaaaatgggaaagaaatgggtccatctctga 545  
Db 249 CAGACGTTCTGAAAAAGACACACAGATCTCTGGGAAGGAACATCCCTATTTCTGG 190  
Qy 546 acaagtacagaatgaaatcaaatcaatgaagaaagaatgagtgatgtatgtatt 605  
Db 189 CATGGCACTGATTAGAAACAGTCAAAAAAAGAAAAAAGAAATATCCCTATTT 130  
Qy 606 taacaaaaa 614  
Db 129 TAACCACAA 121  
RESULT 14  
US-08-253-503-1/c  
; Sequence 1, Application US/08253503  
; Patent No. 5589334  
; GENERAL INFORMATION:  
; APPLICANT: Coullie, Pierre  
; APPLICANT: Boon-Falleur, Thierry  
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE WHICH CODES  
; TITLE OF INVENTION: FOR A TUMOR REJECTION ANTIGEN PRECURSOR  
; TITLE OF INVENTION: WHICH IS PROCESSED TO AN ANTIGEN PRESENTED  
; TITLE OF INVENTION: BY HLA-B\*44, AND USES THEREOF  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Felfe & Lynch  
; STREET: 805 Third Avenue  
; CITY: New York



us-09-833-782-1.rni

Tue Jan 15 09:01:14 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:56:36 ; Search time 63.23 Seconds  
(without alignments)  
408.225 Million cell updates/sec

Title: US-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCLLAVSRRLVGGSR.....FLKREPQKAFILMSRGLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3483	95.0	704	1 NEUL_PIG	Q02038 sus scrofa
2	3475	94.7	704	1 NEUL_RABIT	P42675 oryctolagus
3	3348	91.3	704	1 NEUL_RAT	P42676 rattus norv
4	2330	63.5	686	1 MEPL_PIG	P47788 sus scrofa
5	2330	63.5	686	1 MEPL_HUMAN	P52888 homo sapien
6	2291	62.5	686	1 MEPL_RAT	P24155 rattus norv
7	1085.5	29.6	712	1 PRVD_YEAST	P25375 saccharomyc
8	818.5	22.3	680	1 OPDA_ECOLI	P27298 escherichia
9	814.5	22.2	681	1 OPDA_HAEIN	P44573 haemophilus
10	806	22.0	680	1 OPDA_SALTY	P27237 salmoneilla
11	525	14.3	713	1 PMTP_HUMAN	Q99797 homo sapien
12	519	14.1	680	1 DCP_ECOLI	P24171 escherichia
13	502.5	13.9	680	1 DCP_SALTY	P27236 salmoneilla
14	472	12.9	710	1 PMTP_RAT	Q01992 rattus norv
15	461	12.6	772	1 PMTP_YEAST	P35999 saccharomyc
16	448.5	12.2	762	1 PMTP_SCHPO	Q10415 schizosacch
17	405.5	11.1	775	1 PMTP_SCHCO	P37932 schizophyll
18	145.5	4.0	3911	1 AKA9_HUMAN	Q99996 h a kinase
19	133.5	3.6	590	1 PEFF_BORBU	O51264 borrelia bu
20	133	3.6	1341	1 RPAL_METJA	Q58445 methanococ
21	132	3.6	993	1 SCPL_MOUSE	Q62209 mus musculu
22	132	3.6	1427	1 REST_HUMAN	P30622 homo sapien
23	131.5	3.6	1818	1 T294_HUMAN	O94822 homo sapien
24	129.5	3.5	1539	1 Y373_HUMAN	O15078 homo sapien
25	129.5	3.5	3066	1 POLG_SBMVG	Q90069 s genome po
26	129	3.5	1972	1 MYHB_HUMAN	P35749 homo sapien
27	128	3.5	834	1 YN4_CAEEL	P34537 caenorhabdi
28	125	3.4	591	1 YN48_YEAST	P42846 saccharomyc
29	125	3.4	1235	1 DPOL_PYRHO	O59610 pyrococcus
30	124.5	3.4	5035	1 RYNR_PIG	P16960 sus scrofa
31	123.5	3.4	1087	1 AKA9_RABIT	Q28628 oryctolagus
32	123	3.4	997	1 SCPL_RAT	Q03410 rattus norv
33	122.5	3.3	602	1 PEP2_LACLC	P94876 lactococcus

#### ALIGNMENTS

RESULT	NEUL_PIG	STANDARD;	PRT;	704 AA.
AC	Q02038:			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DE	NEUOLYSIN PRECURSOR (EC 3.4.24.16) (NEUTENSIN ENDOPEPTIDASE)			
DE	(MITOCHONDRIAL OLIGOPEPTIDASE M) (MICROSOMAL ENDOPEPTIDASE) (MEP)			
DE	(SOLUBLE ANGIOTENSIN-BINDING PROTEIN) (SABP) (ENDOPEPTIDASE 24.16).			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Eumalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Heart, and Liver;			
RX	MEDLINE=92388174; PubMed=1517239;			
RA	Sugiura N., Hagiwara H., Hirose S.;			
RT	"Molecular cloning of porcine soluble angiotensin-binding protein.";			
RL	J. Biol. Chem. 267:18067-18072(1992).			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=97326108; PubMed=9182559;			
RA	Kato A., Sugiura N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;			
RT	"Targeting of endopeptidase 24.16 to different subcellular			
RT	compartments by alternative promoter usage.";			
RL	J. Biol. Chem. 272:15313-15322(1997).			
CC	-!- FUNCTION: BINDS ANGIOTENSINS AND ITS ANALOGS.			
CC	-!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE IN NEUTENSIN:			
CC	10-PRO- -TYR-11.			
CC	-!- COFACTOR: BINDS ONE ZINC ION.			
CC	-!- SUBCELLULAR LOCATION: MITOCHONDRIAL AND CYTOPLASMIC			
CC	(BY SIMILARITY).			
CC	-!- TISSUE SPECIFICITY: IS PREDOMINANTLY EXPRESSED IN THE LIVER,			
CC	KIDNEY AND ADRENAL GLAND.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);			
CC	ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D11336; BAA01949.1;			
DR	EMBL; AB000170; BAA19060.1;			
DR	EMBL; AB000425; BAA19104.1;			
DR	EMBL; AB000411; BAA19104.1; JOINED.			
DR	EMBL; AB000414; BAA19104.1; JOINED.			
DR	EMBL; AB000415; BAA19104.1; JOINED.			
DR	EMBL; AB000416; BAA19104.1; JOINED.			
DR	EMBL; AB000417; BAA19104.1; JOINED.			

p35579 homo sapien  
p49454 homo sapien  
q01397 neurospora  
q9usi6 schizosacch  
p11047 homo sapien  
q13439 homo sapien  
p02566 caenorhabdi  
p35748 oryctolagus  
p32380 saccharomyc  
p21817 homo sapien  
p79134 bos taurus  
p06529 bacillus su





```

QY 181 PEQVONEIKSMKRMSELCLIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDGDKYK 240
Db 181 PEEVQNEIKSMKRMSELCLIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDGDKYK 240
QY 241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVKALLGYSTHA 300
Db 241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVKALLGYSTHA 300
QY 301 DFVLEMTAKTSRVTAFDLDLSOKLPLGEAREFIIILNKKKECKDRGFEYDGNKINAWD 360
Db 301 DFVLEMTAKTSRVTAFDLDLSOKLPLGEAREFIIILNKKKECKDRGFEYDGNKINAWD 360
QY 361 LYYMTQTEELKYSIDQEFKEFPIEVTEGLNTYQELLGLSFQMTDAHVNNKSVTL 420
Db 361 LHYMTQTEELKYSIDQEFKEFPIEVTEGLNTYQELLGLSFQMTDAHVNNKSVTL 420
QY 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGRSMMAVAALVYVNSQPV 480
Db 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGRSMMAVAALVYVNSQPV 480
QY 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWVDVD 540
Db 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWVDVD 540
QY 541 SLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600
Db 541 SLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600
QY 601 AKYCSILGVAATPGTNMPATFGLHAGGYDQGYGYLWSEVFSMDMFCFKKEGIMNPE 660
Db 601 ARYCTDILGVAATPGTNMPATFGLHAGGYDQGYGYLWSEVFSMDMFCFKKEGIMNPE 660
QY 661 VGMKYRNLILKPGSLDGDMDLHNFILKREPNOKAFILMSRGLHAP 704
Db 661 VGMKYRNLILRPGSLDGDMDLHNFILKREPNOKAFILMSRGLQAP 704

RESULT 3
NEUL_RAT
ID NEUL_RAT STANDARD; PRT; 704 AA.
AC P42676;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NEULOLYSIN PRECURSOR (EC 3.4.24.16) (NEUROTENSIN ENDOPEPTIDASE)
DE (MITOCHONDRIAL OLIGOPEPTIDASE M) (MICROSOMAL ENDOPEPTIDASE) (MEP).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX MEDLINE=96070836; PubMed=7592986;
RA Dauch P., Vincent J.-P., Checler F.;
RT "Molecular cloning and expression of rat brain endopeptidase
RT 3.4.24.16.";
RL J. Biol. Chem. 270:27266-27271(1995).
RN [2]
RP SEQUENCE OF 38-57, AND CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=95138171; PubMed=7836437;
RA Serizawa A., Dando P.M., Barrett A.J.;
RT "Characterization of a mitochondrial metalloproteinase reveals
RT neurolysin as a homologue of thimet oligopeptidase.";
RL J. Biol. Chem. 270:2092-2098(1995).
CC -!- FUNCTION: HYDROLYSES OLIGOPEPTIDES SUCH AS NEUROTENSIN,
CC BRADYKININ, DYNORPHIN A, ETC.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE IN NEUROTENSIN:
CC 10-PRO-|-IYR-|-I.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INTERMEMBRANE SPACE AND
CC ALSO CYTOPLASMIC.

```

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
 ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X87157; CAA60630.1; -  
 DR MEROPS: M03.002; -  
 DR InterPro: IPR001567; Peptidase\_M3.  
 DR InterPro: IPR000130; Zn\_Metpeptidse.  
 DR Pfam: PF01432; Peptidase\_M3; 1.  
 DR PROSITE: PS001442; ZINC\_PROTEASE; 1.  
 KW Metalloprotease; Hydrolase; Zinc; Mitochondrion; Transit peptide.  
 FT TRANSIT 1 37 MITOCHONDRION.  
 FT CHAIN 38 704 NEUROLYSIN.  
 FT ACT\_SITE 498 498 BY SIMILARITY.  
 FT METAL 497 497 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 501 501 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 504 504 ZINC (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 704 AA; 80253 MW; E33F7967A79343D1 CRC64;  
 -----  
 Query Match 91.3%; Score 3348; DB 1; Length 704;  
 Best Local Similarity 89.7%; Pred. No. 1.8e-199;  
 Matches 630; Conservative 40; Mismatches 32; Indels 0; Gaps 0;  
 -----  
 QY 1 MATRCILAVRSRRVGGSRILLRMTLGRVMSPLQAMSSYTVAGRVNLRWDLSPQLIKTR 60  
 Db 1 MITCLSLTLRGLHRAGGSRLLQLTLTGKELASPLQAMSSYTAAGRVNLRWDLSPQLIKTR 60  
 QY 61 TEELIVQTKQVYDAVGMGLGIEEVTYENCLOALADVEVKYIVERTMLDFFQHVSSDREVA 120  
 Db 61 TEOLIAQTKQVYDVTGCTIALKEVTYENCLOVLADIEVTYIVERTMLDFFQHVSSDREVA 120  
 QY 121 ASTEADKRLSRFDIEMSRMGDIPERIVHLOETCDLQKIKPEARRYLEKSIKMGKRNGLHL 180  
 Db 121 ASTEADKLSRFDIEMSRMGDVFQIRVHLOETCDLEKIKPEARRYLEKSIKMGKRNGLHL 180  
 QY 181 PEQVONEIKSMKRMSELCLIDFNKNLNEDDTFLVFSKAELGALPDDFIDSLEKTDGDKYK 240  
 Db 181 SEHIREIKSMKRMSELCLIDFNKNLNEDDTSLVFSKAELGALPDDFIDSLEKTDGDKYK 240  
 QY 241 ITLKYPHYFPVVKKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVKALLGYSTHA 300  
 Db 241 VTLKYPHYFPVVKKCCVPETRRKMEMAFNTRCKQENTAILQQLPLRAQVAKLGYNTHA 300  
 QY 301 DFVLEMTAKTSRVTAFDLDLSOKLPLGEAREFIIILNKKKECKDRGFEYDGNKINAWD 360  
 Db 301 DFVLELNTAKTSRVAFAFDLDLSOKLPLGEAREFIIILNKKKECKDRGFEYDGNKINAWD 360  
 QY 361 LYYMTQTEELKYSIDQEFKEFPIEVTEGLNTYQELLGLSFQMTDAHVNNKSVTL 420  
 Db 361 LHYMTQTEELKYSVDQESLKEFPIEVTEGLLSIYQELLGLSFQVQPDAAHVNNKSVSL 420  
 QY 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGRSMMAVAALVYVNSQPV 480  
 Db 421 YTVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCGLLPDGRSMMAVAALVYVNSQPV 480  
 QY 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWVDVD 540  
 Db 481 AGRPSLLRHDEVRTYFHEFGVHMHIQCAOTDFARFSGTNVETDFVEVPQMLNWWVDVD 540  
 QY 541 SLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600  
 Db 541 SLRRLSKHYKDGHPITDDELLEKLVASRLVNTGLTLRQIVLSKVQDQSLHTNTSLDAASEY 600  
 QY 601 AKYCSILGVAATPGTNMPATFGLHAGGYDQGYGYLWSEVFSMDMFCFKKEGIMNPE 660  
 Db 601 AKYCSILGVAATPGTNMPATFGLHAGGYDQGYGYLWSEVFSMDMFCFKKEGIMNPE 660

Db 601 AKYTEILGVAATPGTNMPATFGHLAGGYDQYGYUWSEVFSMDMFHSCFKKKGIMNPE 660  
QY 661 VGMKYNLILKPGSLDGMHLNFKREPQKAFILMSRGLH 702  
DB 661 VGMKYNLILKPGSLDGMHLNFKREPQKAFILMSRGLH 702  
RESULT 4  
MEPD\_PIG STANDARD; PRT; 686 AA.  
AC P47786;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDOPEPTIDASE 24.15).  
GN THOPI.  
OS Sus scrofa (pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA MEDLINE=94222057; PubMed=8168506;  
RX Kato A., Sugliura N., Hagiwara H., Hirose S.;  
RT "Cloning, amino acid sequence and tissue distribution of porcine  
RT thimet oligopeptidase. A comparison with soluble angiotensin-binding  
RT protein.";  
RL Eur. J. Biochem. 221:159-165 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97326108; PubMed=9182559;  
RA Kato A., Sugliura N., Saruta Y., Hosoiri T., Yasue H., Hirose S.;  
RT "Targeting of endopeptidase 24.16 to different subcellular  
RT compartments by alternative promoter usage.";  
RT J. Biol. Chem. 272:15313-15322 (1997).  
CC -1- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
CC DEGRADATION  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-1-ALA BOND  
CC IN SUBSTRATES CONTAINING -PHE-PHE-1-ALA-ALA-PHE- AND SIMILAR  
CC SEQUENCES.  
CC -1- COFACTOR: BINDS ONE ZINC ION.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D21871; BAA04882.1; --  
DR EMBL; AB000438; BAA19107.1; --  
DR EMBL; AB000426; BAA19107.1; JOINED.  
DR EMBL; AB000427; BAA19107.1; JOINED.  
DR EMBL; AB000428; BAA19107.1; JOINED.  
DR EMBL; AB000429; BAA19107.1; JOINED.  
DR EMBL; AB000430; BAA19107.1; JOINED.  
DR EMBL; AB000431; BAA19107.1; JOINED.  
DR EMBL; AB000432; BAA19107.1; JOINED.  
DR EMBL; AB000433; BAA19107.1; JOINED.  
DR EMBL; AB000434; BAA19107.1; JOINED.  
DR EMBL; AB000435; BAA19107.1; JOINED.  
DR EMBL; AB000436; BAA19107.1; JOINED.  
DR EMBL; AB000437; BAA19107.1; JOINED.  
DR MEROPS; M03.001; --

DR InterPro; IPR001567; Peptidase\_M3.  
DR InterPro; IPR000130; Zn\_Mrpeptidse.  
DR Pfam; PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Metalloprotease; Hydrolase; zinc.  
FT INIT\_MET 0 BY SIMILARITY.  
FT METAL 472 472 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 473 473 BY SIMILARITY.  
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 686 AA; 78013 MW; AA6006223047580C CRC64;  
Query Match 63.5%; Score 2330; DB 1; Length 686;  
Best Local Similarity 64.0%; Pred. No. 1.3e-136;  
Matches 431; Conservative 104; Mismatches 130; Indels 8; Gaps 1;  
QY 29 EVMSPLQAMSSYTVAGRNVLWDLSPOIKRTREELIVQTVQVYDVGMLGIEEYVENC 88  
DB 12 DVAAPCSAY-----NYLRWDLSSAQIGELTTELEQTKRVYDRVGTQQLQDVSYENT 63  
QY 89 LQALADVEVKYIVERTMLDPHQHVSDEKVEVRAASTEADKRLSRFEDIEMSRGDIIFERIVH 148  
DB 64 LKALADVEVSYTVQVRNILDLPQHVSPCKDIRTASTADKILSEEDVEMSRQDYQRIWV 123  
QY 149 LQETCDLGLKIKPEARRYLEKSIMKGRNGLHLPQVONEIKSMKRMSELICIDFNKNLNE 208  
DB 124 LQEVQKQDSLRLPEAAVYLERLIKLRNGLHLPKETQEKIKSIKKLSLLCIDFNKNLNE 183  
QY 209 DDTFLVFSKAEALGALPDDFIDSLKTDKDKYKITLKYPHYPPVMKKCCIPETRRMENAF 268  
DB 184 DTTFLPVTREELGGLPEDFLNSLEKTEDEKLUKVTLYKPHYFLPKKCHVPETRRKVEAF 243  
QY 269 NTRCKEENTILQOLLPLRTKVAKLGYSTHADFVLENTAKTSRVTAFLLDLSQKLKP 328  
DB 244 NCRCKEENCALRELRLRACKSSILGFEHADVLENNAKTSQVATFELDLAQLKP 303  
QY 329 LGEAREEFILNKKKECKDRGFEDGKINAWDLIYYMTQTEELKYSIDQEFKYEFPLEV 388  
DB 304 LGEQERAVILELKAECTKRGDLDFGRINAWDMRYMNVQVEETRYVDQNLLEKFFPMQV 363  
QY 389 VTEGLLNTYQELLGLSPEQMTDAHWNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNH 448  
DB 364 VTRGLGIYQELLGLTFLHEGAAVHEDVMLYSVRDAAKGVIGKFYLDLYPREGKYGH 423  
QY 449 AACFGLOPGCLLPDGRMMVAALVNF SQPVAGRPSSLRRHDEVRTYFHEFGHVMHQICA 508  
DB 424 AACFGLOPGCLRPQDGRQIAIAAMVANFTKPTDPAPSLQLQHDEVETYEFGHVMHQLCS 483  
QY 509 QTDFAFSGTNVETDVEVPQSMLENVWVDVLSRLSKHYKDGSPDIADLLEKLVASRL 568  
DB 484 QAEEFAMFSGTHVERDFVEAPSQMLENWWAEPLLRMSQHYRTGSAIPEOLELLEKLSRQ 543  
QY 569 VNTGLLTQLQVLKQDQSLHTNTSLDAASEYAKVCSILGVAATPGTNMPATFGHLAGG 628  
DB 544 ANTGLFNLRQVLAKVDQALHTQTAADPAEYARLCQILGVPATPGTNMPATFGHLAGG 603  
QY 629 YDGQYGYLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNLILKPGSLDGMHLNFKLR 688  
DB 604 YDAQYGYLWSEVYSADMFHTREFQEGILSGKVGMDYRSCILRPGGSEDAVYMLKFLGR 663  
QY 689 EPNOKAFILMSRGL 701  
DB 664 DPKQDAFLLSKGL 676  
RESULT 5  
MEPD\_HUMAN STANDARD; PRT; 688 AA.  
ID MEPD\_HUMAN  
AC P52888;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDOPEPTIDASE 24.15) (MP78).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95367027; PubMed=7639763;  
 RA Thompson A., Huber G., Walther P.;  
 RT "Cloning and functional expression of a metalloendopeptidase from  
 human brain with the ability to cleave a beta-APP substrate  
 peptide.";  
 RL Biochem. Biophys. Res. Commun. 213:66-73(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Taylor G.R., Otulakowski G., Lau C.Y., Munroe D.G.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
 CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
 CC DEGRADATION. ABLE TO DEGRADE THE BETA-AMYLOID PRECURSOR PROTEIN  
 CC AND GENERATE AMYLOIDGENIC FRAGMENTS.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-|-ALA BOND  
 CC IN SUBSTRATES CONTAINING -PHE-PHE-|-ALA-ALA-PHE- AND SIMILAR  
 CC SEQUENCES.  
 CC -!- COFACTOR: BINDS ONE ZINC ION.  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
 CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL; 250115; CA90477.1; -;  
 DR EMBL; 029366; AAA82607.1; -;  
 DR MEMOPS; M03.001; -;  
 DR MIM; 601117; -;  
 DR InterPro; IPR001567; Peptidase M3.  
 DR InterPro; IPR000130; Zn\_MTPeptidse.  
 DR Pfam; PF01432; Peptidase\_M3; 1.  
 DR PROSITE; P500142; ZINC\_PROTEASE; 1.  
 KW Metalloprotease; Hydrolase; Zinc.  
 FT INIT\_MET 0 BY SIMILARITY.  
 FT METAL 472 472 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 473 473 BY SIMILARITY.  
 FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
 SQ SEQUENCE 688 AA; 78708 MW; 25ACF9139CD7EDF CRC64;

Query Match 63.5%; Score 2330; DB 1; Length 688;  
 Best Local Similarity 64.9%; Pred. No. 1.3e-136;  
 Matches 426; Conservative 97; Mismatches 133; Indels 0; Gaps 0;

QY 46 NVLRWDLSPQIKTRTEELIVOTKQYDVGMLGTEEVYENCLOALADVVKYIVERTM 105  
 DB 21 NDLRWLSAQAEERTELEIQTKRYDQVGTQEFEDSYESTLKALADVETVTVQRM 80  
 QY 106 LDFPQHVSSDKVEVRAASTADKRLSRFDIEMSRGDIPIERIVHLOETCDLGKIKPEARY 165  
 DB 81 LDFPQHVSPKDIRTASTADKRLSEEDVEMSRREDVQRIVWLQEKVKDSLRPEARY 140  
 QY 166 LKSKTMGRNGLHLPEQVQNEIKSKRMSELCIDFNKNLNEDDTFLVFSKAEGLALPD 225  
 DB 141 LERLIKLRNGLHLPRTEQENIKRIKKKLSLCLIDFNKNLNEDDTFLVFTLQELGGLPE 200

QY 226 DFIDSLEKTDKDKYKITLKYPHYFFVMKKCCIPETRRRMMEMAFNTRCKEENTIIQQQLP 285  
 DB 201 DFLNSLEKMEDCKLKVTLKYPHYFPLLLKKCHVPETRRKVEEAFNCRCKEENCAILLKELVT 260  
 QY 286 LRTKVALLLGYSTHADFLVLEMTAKSTSRVTAFLDLSOKLAPLGEAREFRLNLNKKKEC 345  
 DB 261 LRAQKSRLLGFTHADYVLENNMAKTSQVATFVLDLAQKLPGLGQERAVILELKRAC 320  
 QY 346 KDRGEYDGGKINAWDLYYMTQTEELKYSIDOEFLEKPEPIEVVTEGLLNTYOELLGLSF 405  
 DB 321 ERGLPFDGRIIRANDMYNNQVEETRYCDQNLKEYFPVQVYVHGLLGIQELLGLAF 380  
 QY 406 EQMTDAHVNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAACFGLQPCCLPDGSR 465  
 DB 381 HHEGASAWHEDVRLYTARDAASGEVVGKGYLDLYPREGKYGHAACFGLQPCCLRDGSR 440  
 QY 466 MNAVAAALVYVNSOPVAGRPSSLLRHDEVRYIFHEFGVHMHOICAOITDFARFSGTNETDFV 525  
 DB 441 QIAIAAMVANFTKPTADAPSLQHDDEVETFFHEFGVHMQLCSQAEFAMFSGTHVERDFV 500  
 QY 526 EYPSOMLENWVDVSLRSLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKYD 585  
 DB 501 EAPSOMLENWVWEQEPPLLRMSRHYRTGSAVPRELLEKLTESQANTGLNQLQIVLAKVD 560  
 QY 586 QSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAGGYDQGYGYLWSEYFSMD 645  
 DB 561 QALHTQTDADPAEEYARLQCEILGVPATPGTNMPATFGHLAGGYDQGYGYLWSEYFSMD 620  
 QY 646 MYSYCFKKEGIMNPEVGMKYRNILKPGSLDGMMLHNFLEKREPQKAFLSRGL 701  
 DB 621 MFHTRFKQEGVLNKGMDYRSCILLRPGGSEDASAMLRFLRFLGRDPKQDAFLLSKGL 676

RESULT 6  
 MEPD\_RAT STANDARD; PRT; 686 AA.  
 ID P24155;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE THIMET OLIGOPEPTIDASE (EC 3.4.24.15) (ENDO-OLIGOPEPTIDASE A)  
 DE (ENDOPEPTIDASE 24.15) (PE-PEPTIDASE) (SOLUBLE METALLO-ENDOPEPTIDASE).  
 GN THOP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Testis;  
 RX MEDLINE=91084500; PubMed=2261476;  
 RA Pierotti A., Dong K.-W., Glucksman M.J., Orłowski M., Roberts J.L.;  
 RT "Molecular cloning and primary structure of rat testes  
 metalloendopeptidase EC 3.4.24.15.";  
 RL Biochemistry 29:10323-10329(1990).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=94029935; PubMed=8216239;  
 RA McKie N., Dando P.M., Rawlings N.D., Barrett A.J.;  
 RT "Thimet oligopeptidase: similarity to 'soluble angiotensin II-binding  
 protein' and some corrections to the published amino acid sequence of  
 the rat testis enzyme.";  
 RL Biochem. J. 295:57-60(1993).  
 CC -!- FUNCTION: INVOLVED IN THE METABOLISM OF NEUROPEPTIDES UNDER 20  
 CC AMINO ACID RESIDUES LONG. INVOLVED IN CYTOPLASMIC PEPTIDE  
 CC DEGRADATION.  
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF THE PHE-|-ALA BOND  
 CC IN SUBSTRATES CONTAINING -PHE-PHE-|-ALA-ALA-PHE- AND SIMILAR  
 CC SEQUENCES.  
 CC -!- COFACTOR: BINDS ONE ZINC ION.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- TISSUE SPECIFICITY: IS EXPRESSED ABUNDANTLY IN THE TESTIS.

IT IS ALSO FOUND IN THE LIVER, LUNG AND KIDNEY.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC EMBL; M61142; AAA41586.1; -;  
CC PIR; A36165; HYRTTH.  
CC PIR; S38760; S38760.  
CC MEROPS; M03.001; -;  
CC InterPro; IPR001567; Peptidase\_M3.  
CC InterPro; IPR000130; Zn\_MTPeptidse.  
CC Pfam; PF01432; Peptidase\_M3; 1.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC KW Metalloprotease; Hydrolase; Zinc.  
CC INIT\_MET 0  
CC METAL\_ 472 472 ZINC (CATALYTIC) (BY SIMILARITY).  
CC ACT\_SITE 473 473 BY SIMILARITY.  
CC METAL\_ 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
CC METAL\_ 479 479 ZINC (CATALYTIC) (BY SIMILARITY).  
CC SEQUENCE 686 AA; 78183 MW; F50E6290E0B6EA12 CRC64;

Query Match 62.5%; Score 2291; DB 1; Length 686;  
Best Local Similarity 64.3%; Pred. No. 3.3e-134;  
Matches 422; Conservative 98; Mismatches 136; Indels 0; Gaps 0;

QY 46 NVLRWLSPEOIKTEFTLIVQTVQVDAVGMGLTEFTYENCLOALADVVKYIVERTM 105  
DB 21 NHLRWLSAQOIRALTTLQIQTCVDRVGAQDFEDVSFESTLKALADVETTVQVRI 80  
QY 106 LDFPHVSDSDEVRAASTADKRLSRFDIEMSRGDIPIERVHJQETCDLQKIRPEARRY 165  
DB 81 LDFPHVSPNKDIRAASTADKRLSEFVEMSRQDVQYRVVWLQETPKDLSKPEARRY 140  
QY 166 LKSTKMKCRNGLHPEQVQNEIKSMKRMSELCLDFNKNLNETDFLVFSKALGALPD 225  
DB 141 LERLKLGRNGLHLPDQTPQKIRNKKLLCLDFNKNLNETDFLPFTFRELGLGLPE 200  
QY 226 DFDSLEKTDKDKYITLKYPHYPPVMKCCIPETRRMEMAFNTRCKEENTIILOQLLP 285  
DB 201 DFLNLEKTEDEGKLVTKLYPHYPPVLLKCHVPETRRLLLEAFNCRCKEENAILKELVS 260  
QY 286 LRTKVALKLGYSTHADFLVLENTAKTSRVTAFLDLDSQKLKPLGEAREFILLNKKEC 345  
DB 261 LRAQSNLGLFRTHADYVLENNMAKTSQTATFDELARKLPLGEQERAVILELKEAS 320  
QY 346 KDRGEYDCKINANDLYYIMTQTEELKYSIDQEFKVEFPVEVTEGLLNTYQELLGLSF 405  
DB 321 AKRGLPFDGRHIAHMDRYIMQVEDSRYVDQNLKKEYFPQVTVRGLATYQELLGLTF 380  
QY 406 EQMTDAHVNKSVTLTYVKDKATGVLGQFYLDLYPRGKYNHAAFCGLQPCGLLPDQSR 465  
DB 381 TLEBGAHAWHEDVRLYSVRDAASGEIEGKPYLDLYPRGKYGHAACFGLQPCGLRQDSR 440  
QY 466 MMAVALVYVNSQPVAGRSLLRHDEVITYFHEFGHVMHOICAPDFARFSGTNNVETDFV 525  
DB 441 QLAIAAMVANFTKPTPDVPSLLQHDVEVITYFHEFGHVMHQLCSQAFAFSGTHVERDFV 500  
QY 526 EVPQSMLENVWDVDSLRRLSKHYKDGSPIDADDLLEKLVASRLVNTGLTLRQIVLSKVD 585  
DB 501 EAPQSMLENVWEKEPLMRMSQHYRTGGEAPDELLKLIKRSQAGNAGLNFRIQIVLAKVD 560  
QY 586 QSLHTNTSLDAASEYAKYCSITGLGVAATPGTNMPATFGHLAGGYDQYGYLWSEVFSMD 645  
DB 561 QVLFHQTDVDAEYARLCQELLGVLPATPGTNMPATFGHLAGGYDAQYGYLWSEVFSMD 620

QY 646 MFYSCFKKEGIMNPEVGMKYNRLILKPGGSLDGMMLHNFLLKREPNOKAFILMSRGL 701  
DB 621 MFHTRFQEGVLSPKVGMDYRTSLRPGGSEDASTMLKQFLGRDPKQDAFLLSKGL 676

RESULT 7  
ID PRD\_YEAST STANDARD; PRT; 712 AA.  
AC P25375;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE SACCCHAROLYSIN (EC 3.4.24.37) (PROTEASE D) (PROTEINASE YSCD)  
GN (OLIGOPEPTIDASE YSCD).  
GN PRD1 OR YCL057W OR YCL57W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RX SEQUENCE FROM N.A.  
RX MEDLINE=94139744; PubMed=8307027;  
RA Buechler M., Tislar U., Wolf D.H.;  
RT "proteinase yscd (oligopeptidase yscd). Structure, function and  
RT relationship of the yeast enzyme with mammalian thimet oligopeptidase  
RT (metalloendopeptidase, EP 24.15).";  
RL Eur. J. Biochem. 219:627-639(1994).  
RN [2]  
RP SEQUENCE OF 1-663 FROM N.A.  
RA Alexandraki D., Galland P., Thireos G., Tzermia M.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 663-712 FROM N.A.

RA Voet M., Volckaert G.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COULD BE INVOLVED IN LATE STAGE OF PROTEIN DEGRADATION.  
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF PRO-|-PHE AND ALA-|-ALA BONDS.  
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC EMBL; X76504; CAA54039.1; -;  
CC EMBL; X59720; CAA42388.1; -;  
CC PIR; S19387; S19387.  
CC MEROPS; M03.003; -;  
CC SGD; S0000562; PRD1.  
CC InterPro; IPR001567; Peptidase\_M3.  
CC InterPro; IPR000130; Zn\_MTPeptidse.  
CC Pfam; PF01432; Peptidase\_M3; 1.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC Hydrolase; Metalloprotease; Zinc.  
CC METAL\_ 501 501 ZINC (CATALYTIC) (BY SIMILARITY).  
CC ACT\_SITE 502 502 BY SIMILARITY.  
CC METAL\_ 505 505 ZINC (CATALYTIC) (BY SIMILARITY).  
CC METAL\_ 508 508 ZINC (CATALYTIC) (BY SIMILARITY).  
CC SEQUENCE 712 AA; 81933 MW; 340910B7FDAFBE37 CRC64;

Query Match 29.6%; Score 1085.5; DB 1; Length 712;  
Best Local Similarity 35.2%; Pred. No. 9.5e-60;  
Matches 251; Conservative 130; Mismatches 295; Indels 37; Gaps 13;

QY 20 ILLRWTLGREVNSPL-----QAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVQTKQ 70  
DB 21 ILLRWTLGREVNSPL-----QAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVQTKQ 70





AC P27237;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE OLIGOPEPTIDASE A (EC 3.4.24.70).  
GN PRLC OR OPDA OR OPTA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-18.  
RX MEDLINE=92165739; PubMed=1537805;  
RA Conlin C.A., Miller C.G.;  
RT "Cloning and nucleotide sequence of opda, the gene encoding  
oligopeptidase A in Salmonella typhimurium.";  
RL J. Bacteriol. 174:1631-1640(1992).  
CC -!- FUNCTION: MAY PLAY A SPECIFIC ROLE IN THE DEGRADATION OF SIGNAL  
PEPTIDES AFTER THEY ARE RELEASED FROM PRECURSOR FORMS OF SECRETED  
PROTEINS. CAN CLEAVE N-ACETYL-L-ALA4.  
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF OLIGOPEPTIDES, WITH BROAD  
SPECIFICITY. GLY OR ALA COMMONLY OCCUR AS P1 OR P1' RESIDUES, BUT  
MORE DISTANT RESIDUES ARE ALSO IMPORTANT, AS IS SHOWN BY THE FACT  
THAT 2-GLY-PRO-GLY-1-GLY-PRO-ALA IS CLEAVED, BUT NOT 2-(GLY)5.  
CC -!- COFACTOR: BINDS ONE ZINC ION.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; M84574; AAA27172.1; -  
DR PIR; A42298; A42298.  
DR MEROPS; M03.004; -  
DR StyGene; SG10311; opda.  
DR InterPro; IPR001567; Peptidase\_M3.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF01432; Peptidase\_M3; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc.  
FT METAL 469 469 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 470 470 BY SIMILARITY.  
FT METAL 473 473 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).  
SQ SEQUENCE 680 AA; 76943 MW; CA0F11C80PE13F89 CRC64;  
  
Query Match 22.0%; Score 806; DB 1; Length 680;  
Best Local Similarity 30.4%; Pred. No. 1.6e-42;  
Matches 206; Conservative 129; Mismatches 300; Indels 42; Gaps 17;  
  
QY 52 LSPEQIKTRTEELIVTKQYDAVGMGLTEETVENCLOALADVEKVIYVERTMLDFPOH 111  
Db 17 IRPEHVVPATKALACRAAEVGVAHG-APYSWENLCQPLA--EADDVLGRIFS-I-SH 72  
QY 112 VSSDK---EVRAASTADKRLSRFDIEMSRMGDIFERIVHLQETCDLGKIKPPEARYLEK 168  
Db 73 LNSVKNSPELREAYEOTPLLSSEYSTWVGHEGLYNAYRDLRSDGHYATINTAQKKAVDN 132  
QY 169 SIKMGKRNGLHLPQEQVQNEIKSMKMSLCTDFKNLNEDD---TFLVFSKAEALGALPD 225  
Db 133 ALRDFELSGIGLPKQKQRYGEIATRLSELNGQYSNNVLDTMGWTKLITDEAELAGMPE 192  
QY 226 DEIDS---LEKTDKDKYKITLKYHYFPVMMKCCIPETRRRMEAFNTRC-----K 273  
Db 193 SALAAKAQAEKQGGYLLTLDIPSLFVMTYCDNQALREEMRAYSTRASQDQGNAGK 252  
QY 274 EENTIILQOLLPLRTKVAKLGLGYSTHADEVLEMNTAKSTSRVTAFLDLDLSQKLKPLGEAE 333

Db 253 WDNSPVMEEILALRHELAQLLGFENYAHESLATKMAENPQQVLDFTDLAKRARPOQGEKE 312  
QY 334 REFILNLKKKECKDRGFEDYDGINAWDLYYMTQTTELKYSIDQEFKXYFPIEVTEGL 393  
Db 313 LAQLRAFAKAEF---GVE---ELQPWDIAIYSEKQHLIYSISDEQLRYPFENKAVNGL 366  
QY 394 LNTYQELLGLSPQMTDAHWNKSVLYTVKDKATCEVLGQFVLDLYPREGKYNHA---A 450  
Db 367 FEVVKRIYGITAKERTDQVWHPVEVRFELYDE--NNELRGSFYLDLYAREHKGGAMDD 425  
QY 451 CFGLQPGCLLPDGSRMMAVAALVNFSPQVAGRPSLLRHDEVRTEYFEHGHVHQCAQT 510  
Db 426 CVGQMRKA--DGTLOKPVAYLTNCFNRPVNGKPAFTHTDEVITLFEHFGHGLHMLTRI 482  
QY 511 DPARFSG-TNVEDTFVEVPSOMLENVWVDLSRLRSKHYKDGSPDADDLLEKLVASRLV 569  
Db 483 ETAGVSCISCVPMDAVELPSQFMENWCPEALAFSTGHYETGEPLPKELLDKMLAAKNY 542  
QY 570 NTGLLTRLQVLVLSKVDQSLHTNTSLDAASEYAKYCSSEI-LGVAATPCT---NMPATFGHL 625  
Db 543 QAALFLRLQLEGLDFRLHAENPQOGAKILETLEIKKQVAVVSPTWGRPHAFSHI 602  
QY 626 -AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPVGMKYRNLILKPGGSLDGMDLHN 684  
Db 603 FAGGYAAGYVYLWADVLAADA-YSRFEEBEGFNRETGQSFLDNLTGRGSEEPMELEKFR 661  
QY 685 FLKREPQKRAFLMSRGL 701  
Db 662 FRGREPQLDAMLEHYGI 678  
  
RESULT 11  
PMIP\_HUMAN  
ID PMIP\_HUMAN STANDARD; PRT; 713 AA.  
AC Q99797;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).  
GN MIP-OR MIP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=97230465; PubMed=9073519;  
RA Chew A., Buck E.A., Peretz S., Sirugo G., Rinaldo P., Isaya G.;  
RT "Cloning, expression, and chromosomal assignment of the human  
mitochondrial intermediate peptidase gene (MIP).";  
RL Genomics 40:493-496(1997).  
CC -!- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO  
THEIR MATURE SIZE.  
CC -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND  
STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE  
MITOCHONDRION.  
CC -!- ENZYME REGULATION: ACTIVITY IS DIVALENT CATION-DEPENDENT. IT IS  
STIMULATED BY MANGANESE, MAGNESIUM OR CALCIUM IONS AND REVERSIBLY  
INHIBITED BY ZINC, COBALT, AND IRON (BY SIMILARITY).  
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);  
ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----







```

DR PIR; A49931; A49931.
DR MEROPS; M03.005; -.
DR EcoGene; EG10212; dcp.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Dipeptidase;
KW Calcium; Complete proteome.
FT INIT_MET 0
FT METAL 469 469 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 470 470 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 473 473 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).
FT CONFLICT 138 139 IH -> LL (IN REF. 1).
SQ SEQUENCE 680 AA; 77384 MW; 04B9DCC03E15893A CRC64;

Query Match 14.1%; Score 519; DB 1; Length 680;
Best Local Similarity 25.7%; Pred. No. 8.6e-25;
Matches 157; Conservative 120; Mismatches 299; Indels 34; Gaps 12;

QY 114 SKEVRASTEADKRLSRDIEMSGDIFERIVHQETCDLGKIKPEARVYLEKSIKMG 173
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
83 TNDLQRLDEQFSAEALAEALNDIYLNGELFARVDVWQRRESGLDSEIRLVEIHQRF 142
QY 174 KRNLHLPQOVONEIKSMKRMSELICIDFNKML--NEDDTFLVFSKALGALPDDFI-- 228
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 VLAKAKLAQAKAKLVLTAEATTSQFNORLLANKSGGLVNDIAGLQMSQEIAL 202
QY 229 ---DSLEKDDDKYKITLKYPHYFPVKKKCCIPETRRMEMAFNTRCKE---ENTIILO 281
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 AAFAAREKGLDNKWLPIPLNTQQPALAEWRDRATEKLFIAGTWRAEKNDANDTRAIQ 262
QY 282 QLLPLRTKVLKLGSTHADFLVLENNATKSTSRVTAFLDLSQKLPGLGEAREFILTUK 341
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
263 RLVEIRAQOATLILGFHYAAKIAIDOMAKTPEAALNFMREIVPAARQASDELASIQAVI 322
QY 342 KKECKDRGFEYDGIKINAWDLYXYMTQTEELKYSIDOEFLKEYFPPIE-VVTEGLLTYQEL 400
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 DKQ--OGGF---SAQPDWAFYAEQVRREKFDLDLDAQLKPYFELNTVLNEGVTANQL 376
QY 401 LGLSFEQMTDAHWKNSVTLTVKDKATGEVLGQFVLDLYPREGKYNHAAFCGLOPGCLL 460
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
377 FGKIFVEREDIPYHPDVRVWEIFDH-NGVGLALFGVDFEARDKSG-----GAWMGNFV 430
QY 461 PDG--SRMAVAALVNFQSVAGRPSRLRHDEVRTYFHEFGVHMHQICAQTDFAFSGT 518
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 EQSTLNKTHPIVYVNCYOKPAAGEPALLLDVITLPHFEGHTLGLFARQRYATLSGT 490
QY 519 NVETDFEVPQMLENVWDVDSRLSKHYKDGSPDIADDLLEKLVASRLVNTGLTLRQ 578
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 NTRDVEFPQSLNEHWATHPQVFARYAHYOSGAMPDELOQKMNASLFNKGYESSEL 550
QY 579 IVLKSKVDQSLHNTSLDAAS-----EYAKYCSFELGVAATPGTNPATGHL-AGGYDQ 632
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
551 LSAALDMRWHLCEANEAMQDVDDFELRALVAENMDLPAIPRYSRYFAHIFGGGYAG 610
QY 633 YGYLWSEVFSMDMFFYCPKKGIMNPEVGMKYRNILKPGGSLDGMDLHNLKREPNO 692
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
611 YYAYLWTOMLA-DDGYQWFEVQGLTRENGRLFREALLSRGNSDEDLERYQWRGKAPKI 669
QY 693 KAFLSRGLH 702
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
670 MPMLQHRGLN 679

RESULT 13
DCP_SALTY
ID DCP_SALTY STANDARD; PRT; 680 AA.
AC P27236;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEPTIDYL-DIPEPTIDASE DCP (EC 3.4.15.5) (DIPEPTIDYL CARBOXYPEPTIDASE).
GN DCP.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-12.
RX MEDLINE=92165738; PubMed=1537804;
RA Miller C.G., Hamilton S.;
RT "Cloning and nucleotide sequence of the Salmonella typhimurium dcp
   " gene encoding dipeptidyl carboxypeptidase.";
RL J. Bacteriol. 174:1626-1630(1992).
CC -!- FUNCTION: REMOVES DIPEPTIDES FROM THE C-TERMINI OF N-BLOCKED
   TRIPEPTIDES, TETRAPEPTIDES AND LARGER PEPTIDES.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF UNBLOCKED, C-TERMINAL DIPEPTIDES
   FROM OLIGOPEPTIDES, WITH BROAD SPECIFICITY. DOES NOT HYDROLYSE
   BONDS IN WHICH PL' IS PRO, OR BOTH PL AND PL' ARE GLY.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
   ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
EMBL; M84575; AAA27055.1; -.
DR PIR; A42297; A42297.
DR MEROPS; M03.005; -.
DR StyGene; SG10083; dcp.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_Mtpeptidase.
DR Pfam; PF01432; Peptidase_M3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Dipeptidase.
FT ACT_SITE 469 469 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 473 473 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 476 476 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 680 AA; 77261 MW; DDFD89B47B660B45 CRC64;

Query Match 13.7%; Score 502.5; DB 1; Length 680;
Best Local Similarity 26.8%; Pred. No. 9e-24;
Matches 190; Conservative 115; Mismatches 302; Indels 103; Gaps 24;

QY 61 TEEILVQTKQVDAVGMGLGIEEVTYENCILQ---ALADVEVKYIVE-----RTMLDFP 109
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
3 TNPLLDQSLPQAPRFDRIKDCYRPAFDEGVQRKRVIEAIVNHVPAAPDFNTLLALE 62
QY 110 QH---VSSDKVRASTEAKDKRLSRFDIEMS-----MRGDI FERIVHL-QET 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 QSGALLSRVTSVFFAMTAHTNDELQRLDEAFSAELAAALSNDIYLSALFAKVDVWQOR 122
QY 153 CDLGKIKPEARVYLEKSIKMKRNGHLHLPQOVONEIKSMKRMSELICIDFNKMLNEDDTF 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123 HSLG-LDDESRLVDVIHORFVLAGAQLAEEDKARLKVLTESATLMSQFNQRL----- 175
QY 213 LVFSKALGALPDDFIDSL-----EKDDDKYKITLKYPHYFPVVMKKCC 256
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
176 LAASKA--GGLAVDDAHCLAGLSPEEMTVAAEAAREKGLEERWFIPLLNTTQOPALATLR 233
QY 257 IPETRRRMEAFNTRCKE-----ENTIILOQLPLRTKVLKLGSTHADFLVLENNATKST 312
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 DRQTRNLFAASWTRAERKGDADHTRAIVQRLVEIRRCQAKLLGFPNYAAWKADQAKTP 293
QY 313 SRVTAFLDLSQKLPGLGEAREFILTUKKCKDRGFEYDGIKINAWDLYXYMTQTEELKYSIDOEFLKEYFPPIE-VVTEGLLTYQEL 367

```

```
Db 294 QAALSF-----MRGIVPPARQRLN-EQAIEQN-----VIDGEGGYTVQAWDMFYAEQ 342
QY 368 TEEKYSIDQEFKEYPIE-VVTEGLNYYQELGLSEFQMTDAHVWNKSVTLTYVKDK 426
Db 343 VRREKALDEAQLKPYFALNTVLOEGFWTANQIFLTFVEREDIPVHPDVRWEIFD- 401
QY 427 ATGEVLGQFYLDLYPREGKYNHAAFCGLQGLLPDGL--SRMMAVAALVNFSPVAGRP 484
Db 402 SDGVGMALFYCDFPARKSG-----CAWGNFVEQSTLNETRPVIYNVCNYQKPVGDQP 456
QY 485 SLRLHDEVRYTFHEFGHVMHQICAQTFARFSGTNVETDFVEVPSQMLENNWVDVSLRR 544
Db 457 ALLLWDVVITLFEHFGHTLGLFAVQRYATLSGTNTPRDFVEFVPSQINEHWASPRVFER 516
QY 545 LSKHYKDGSPADLLLEKLVASRLVNTG-----LLTLQIVL--SKVDQSLHTNT 592
Db 517 YARVHDEGKMPADQERMKRKLNFNGYDMTELLGAALLDMRMHMLEESVAEQSV----- 572
QY 593 SLDAASEYAKYCSITLVAATPGTNMPATFGLH--AGGYDGOYGYGLWSEVFSMDMFYSCF 651
Db 573 --REFEQOALAAEHLDPVAVPPRYSYFAHIFGGGYAAGYAYLWTQMLA-DDGYQWF 628
QY 652 KKEGIMPEVGMKYNLILKPGSLDGMMLHNLKREPNOKAFILMSRGL 701
Db 629 VEQGLTRENGQRFDAIARGNSTDLTLYSAWRGHEPHIDPMLQYRGL 678

RESULT 14
PMIP_RAT ID PMIP_RAT STANDARD; PRT; 710 AA.
AC Q01992;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).
GN MIPEP OR MIP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92390438; PubMed=1518864;
RA Issaya G., Kalousek F., Rosenberg L.E.;
RT "Sequence analysis of rat mitochondrial intermediate peptidase:
RT similarity to zinc metalloproteinases and to a putative yeast
RT homologue.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8317-8321(1992).
RN [2]
RP SEQUENCE OF 34-50 AND 528-542.
RC TISSUE=Liver;
RX MEDLINE=92347312; PubMed=1322290;
RA Kalousek F., Issaya G., Rosenberg L.E.;
RT "Rat liver mitochondrial intermediate peptidase (MIP): purification
RT and initial characterization.";
RL EMBO J. 11:2803-2809(1992).
CC -1- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO
CC THEIR MATURE SIZE.
CC -1- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND
CC STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE
CC MITOCHONDRION.
CC -1- ENZYME REGULATION: ACTIVITY IS DIVALENT CATION-DEPENDENT. IT IS
CC STIMULATED BY MANGANESE, MAGNESIUM OR CALCIUM IONS AND REVERSIBLY
CC INHIBITED BY ZINC, COBALT, AND IRON.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
-----
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M96633; AAA1899.1; -.
DR PIR; A46273; A46273.
DR MEROPS; M03.006; -.
DR InterPro; IPR001567; Peptidase_M3.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF01432; Peptidase_M3.1
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydroxylase; Metalloprotease; Zinc; Transit peptide; Mitochondrion;
KW Magnesium; Manganese; Calcium; Cobalt; Iron.
FT CHAIN 1 33 MITOCHONDRION.
FT METAL 34 710 MITOCHONDRIAL INTERMEDIATE PEPTIDASE.
FT ACT_SITE 492 492 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 493 493 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 496 496 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 499 499 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 710 AA; 80673 MW; 4FC2E6743EA43558 CRC64;

Query Match. 12.9%; Score 472; DB 1; Length 710;
Best Local Similarity 24.4%; Pred. No. 7.3e-22;
Matches 187; Conservative 119; Mismatches 294; Indels 166; Gaps 28;

QY 17 GSRTLLRMTLGRVMSPLQAMSSYTVAGRNVL-----AGSCARSVSTSWSPVGAFAFNKPGQHLWDLLE 52
Db 6 GTRYAYRL-CGRRAAALQGR-----AGSCARSVSTSWSPVGAFAFNKPGQHLWDLLE 59
QY 53 -----SPEQIKTRTEELIVQTKQYDAV-----GMLGIEEYVYENCALQA 93
Db 60 RRGFGVPPELSTPEGFQVAQEALRKTELVERACSTPPGGPQTVLIEDELS--DCLCRVA 117
QY 94 DVEVKYIVERTMLDFPOHVSSDKEVRAASTFADKRLSRFDIEMSMRGDIERIVHLOETC 153
Db 118 D-----LADFVKIGHPEQAFREAQAEACISGTMVEKLTNTNVELYQ---SLOKLL 164
QY 154 D---LGKIKPEARRYLEKSIKMGKRNGLHLPQEQVQNEIKSMKMRKSELCIDFN-KNLNE 208
Db 165 DDKKLMSLDAETRRVAELPFWDFEISGIHLDEB-----KRRA--VDLNVKILDL 213
QY 209 DDTFLVFS---KAELGALPDDF-----ISL--EKTDD---DKYKITLKP 246
Db 214 SSAFLMGTNFPKIQKHLLPEHIQHFEARDGRHLVDGLHAEASDDLVRFAAYKIFL-YP 272
QY 247 HYFPVMKCCIPETRRRMEMAFNTRCKEENTIILOQLPLRTKVKAKLLGYSTHADFVLEM 306
Db 273 N-----ADQLKC-----LEEELSSRDLLANLVGY-----LPF 299
QY 307 NT-----AKSTSRVTAFLDLSQKLAPLGEAREFILNLKKECKDRGFEDGKINAWD 360
Db 300 PTGPPGTAOTPETVMQFLEKLSKELCTERKDFEMQGMKTKLNQ-----NSELAPWD 354
QY 361 LYYMTQTEELKYSIDQEFKEYPIEVVTEGLNYYQELGLSF--EQMTDAHVWNKSV 418
Db 355 PPYYSGVIRAERYNIERSLYCPFLSLGACMEGLNVLENRLGLVTLYAEQPFKEGVWCIDV 414
QY 419 TLYTVKDKATGEVLGQFYLDLYPREGKYNHAAFCGLQGLLPDGLSRMMAVAALVNFSPQ 478
Db 415 RKLAVVHESEG-LLGYIYCDFQFRANKPQQOCHFTIRGRUKEDGSGYQLPVVLMNLPH 473
QY 479 PVAGRPSILLRHDEVRYTFHEFGHVMHQICAQTFARFSGTNVETDFVEVPSQMLENNWVD 538
Db 474 ASRDFPILLPGMMENLFHEGHAMHSMIGTRYQHVHTGTCTPTDFAEVPSSILMEYSND 533
QY 539 VDSLRRLSKHYKDGSPADLLLEKLVASRLVNTGTLRLQIVLSKVDQSLHTNTSLDAS 598
Db 534 YRVVSQFAKHVYTGQPLPKAMVSRCLCESKSKYCAAAEQMLQVFAALDQIYHGQPL----- 589
QY 599 EYAKYCSIL-----GVAATPCTNMPATFGLHAGGYDGOYGYGLWSEVFSMDMFYSC 650
```

Db 590 --KKSTDIILMETEQFYGLPYVDPFAWQLRFSHLV-GYGAKYIYSLMSRAVSMWVWKEC 646

QY 651 FKKEGIMNEVGMKYNRLILKPGGSLDGLDMLHNFILKREPKNQKAF 696

Db 647 FLODP-FNRAAGERYRRELAHGGKREPLMTIQMLQKPCSIDDFV 691

RESULT 15

PMIP\_YEAST STANDARD; PRT; 772 AA.

AC P35999; P51980;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE MITOCHONDRIAL INTERMEDIATE PEPTIDASE PRECURSOR (EC 3.4.24.59) (MIP).

GN OCT1 OR MIP1 OR YKL134C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94309680; PubMed=8035833;

RA Isaya G., Miklos D., Rollins R.A.;

RT "MIP1, a new yeast gene homologous to the rat mitochondrial intermediate peptidase gene, is required for oxidative metabolism in Saccharomyces cerevisiae.";

RT Saccharomycetes cerevisiae.";

RL Mol. Cell. Biol. 14:5603-5616(1994).

RN [2]

RP SEQUENCE FROM N.A.

RA Rad M.R., Xu G., Kirchrath L., Fritz C., Keuchel H., Hollenberg C.P.;

RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.

CC -!- FUNCTION: CLEAVES PROTEINS, IMPORTED INTO THE MITOCHONDRION, TO THEIR NATURAL SIZE. CLEAVES THE NUCLEAR-ENCODED PRECURSORS FOR CYTOCHROME OXIDASE SUBUNIT IV (COXIV) AND THE IRON-SULFUR PROTEIN (FE-S) OF THE BC1 COMPLEX.

CC -!- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL OCTAPEPTIDE AS SECOND STAGE OF PROCESSING OF SOME PROTEINS IMPORTED IN THE MITOCHONDRION.

CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE); ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U10243; AAA21278.1; -

DR EMBL; Z28134; CAAB1975.1; -

DR PIR; S37963; S37963.

DR MEROPS: M03.006; -

DR SGD; S0001617; OCT1.

DR InterPro; IPR001567; Peptidase\_M3.

DR InterPro; IPR001030; Zn\_Mtpeptdse.

DR Pfam; PF01432; Peptidase\_M3; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Zinc; Transit peptide; Mitochondrion.

FT TRANSIT 1 37

FT CHAIN 38 772 MITOCHONDRIAL (POTENTIAL).

FT METAL 558 558 MITOCHONDRIAL INTERMEDIATE PEPTIDASE.

FT ACT\_SITE 559 559 BY SIMILARITY (BY SIMILARITY).

FT METAL 562 562 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 565 565 ZINC (CATALYTIC) (BY SIMILARITY).

FT CONFLICT 343 350 MAKNPDRV -> WQDRCR (IN REF. 2).

FT CONFLICT 694 694 Y -> S (IN REF. 2).

FT CONFLICT 696 702 ATYSYL -> OLITATY (IN REF. 2).

FT CONFLICT 772 AA; 88182 NW; 1C19A0655FAAE7CA CRC64;

QY SEQUENCE

Query Match 12.6%; Score 461; DB 1; Length 772;

Best Local Similarity 25.6%; Pred No. 3.9e-21;

Matches 176; Conservative 127; Mismatches 291; Indels 94; Gaps 29;

QY 53 SPEQIKTRTEELIVQTKQYVDVAVGMLGIE--EVTYENCIALQALADVEKYIVERTMLDFPQ 110

Db 85 SPDGLRKFSQVSLQAQAEILLDKMRNDFSESGKLTITMNLDRSLDLCR-----VIDLCE 138

QY 111 HVSS-----DKEYRAASTEADKRLSRFDIEMSRGDIIFERIVHLEQETCDLGKI---RPEAR 163

Db 139 FIRSTHPDDAFVRAAQ---DCHEQMEFEFNVLTVDV-----SLCNILKSLVNLNPEVS 187

QY 164 RYLE-KSIRMGK-----RNLHLPEQVQNETKSMKRMSELCIDFNKLNLEDDEDFLVF 215

Db 188 SKLSAEELKVGKILLDDPEKSGIYMPDPVREKFIQISQEISLVGQEE---INHTD---Y 240

QY 216 SKAELGALPDFDIDS-----LEKTD---KYKI-TLKYPHYPPVMKKCCIPETRRR 263

Db 241 PGSNSVYKIPCKDLNDSKVSTFLKQLNCKVKGQNYKVPFTFGYAA-ALLKSCENEMVRKK 299

QY 264 MEMAFNTRCKEENTIILOQLPLRTKVKALLGYSTHADFVLEMTAKSTSRVTAFLDLS 323

Db 300 LMTALHS-CSDKQVKRLSHLILKRALANLMHKTSYAEYQLEGKMAKNPKDVQDFILTM 358

QY 324 QKLKPLGEREREFILNKKKECK---DRGFYDVGKINAWDLVYVMTOTEEI--KYSIDQ 377

Db 359 NNTIETANELKFAELKAKDLKPLTNTNDEILKLVRPMDRDYIYGYKYLQNLPSNPN 418

QY 378 EFLKEFFIEVTVTEGLNTYQELLGLSFE-QMTD-AHVWNKSVTLTYVKDKATGEVLGQF 435

Db 419 KEISYFTLGNVIQGLSDFLQIYIGIRLEPAITDEGETNSPDVRLNVISEEG-IIGII 477

QY 436 YLDLYPREKYNHAACF-----GLOPGCLLPDGSR-MMVAVALVNFQ 478

Db 478 YCDLFRNGKTSNPAHFTVCCSRQIYPSETDFTSIQVG-ENPDGTYFQLPVLVSLVNFSP 536

QY 479 -PVAGRPSL--LRHDEVRYVFEHFGVHMOICQATDFARFSGTNVETDFVEVPSQMLENW 535

Db 537 ILTASKSLCFLQISEVETLFEHMGHAMSLGRTMQLNIGTRCATDFVELPSILMEHF 596

QY 536 VMDVDSILRLSKHYKDGSPITADPLLEKLVASRLVNTGLTLRQIVLSKVQDSLHTN--- 591

Db 597 AKDIRILITKIGHYGTETIQADMLOREMKSTNFLQNCETYSQAKMAMLDQSFHDEKII 656

QY 592 --TSLDAASEYAKYKCEILGVAATPGTNMPATPGHLAGGYDQYGYGLMSEVFSMDMFYS 649

Db 657 DIDNFDVVENY-QALERRLKVLVDQSNMCGRFGHLF-GYGATYYSYLVFORTIASKIWA 714

QY 650 CFKKEGIMNPEVGMKYRNLLILKPGGSILD 677

Db 715 LFEDDPYSRKN-GDKFKKHLKLLKWLGLKD 741

Search completed: January 14, 2002, 19:03:34

Job time: 418 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:44:31 ; Search time 64.93 Seconds  
(without alignments)  
825.918 Million cell updates/sec

Title: US-09-833-782-2  
Perfect score: 3668  
Sequence: 1 MIARCLLAVRSLLRRVGSRI.....FLKREPQKAFIMSRGLHAP 704

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3483	95.0	704	2 A43411	soluble angiotensin
2	3475	94.7	704	2 A45985	microsomal endopep
3	2330	63.5	689	1 HYHUTH	thimet oligopeptid
4	2330	63.5	689	1 HYHUTH	thimet oligopeptid
5	2279	62.1	687	1 HYRTHH	thimet oligopeptid
6	1085.5	29.6	712	2 S19387	saccharolysin (EC
7	848	23.1	680	2 C82353	oligopeptidase A V
8	818.5	22.3	680	2 S47718	oligopeptidase A (
9	817.5	22.3	680	2 B86021	oligopeptidase A (
10	814.5	22.2	681	2 C64055	thimet oligopeptid
11	806	22.0	680	2 A42298	thimet oligopeptid
12	794.5	21.7	681	2 E83636	oligopeptidase A P
13	702	19.1	678	2 B81224	oligopeptidase A N
14	694	18.9	678	2 G81996	oligopeptidase A (
15	690.5	18.8	713	2 S76766	hypothetical prote
16	671.5	18.3	685	2 F75370	oligopeptidase A -
17	635.5	17.3	723	2 T45985	oligopeptidase A X
18	627.5	17.1	674	2 F82844	oligopeptidase A X
19	578.5	15.8	716	2 E82620	peptidyl-di-peptida
20	520	14.2	681	2 D85720	dipeptidyl carboxy
21	519	14.1	681	2 E64908	peptidyl-di-peptida
22	502.5	13.7	680	2 A42297	peptidyl-di-peptida
23	495.5	13.5	678	2 T02867	peptidyl dipeptida
24	472	12.9	710	1 A46273	mitochondrial inte
25	448.5	12.2	762	2 T38081	probable mitochond
26	415	11.3	770	2 S37963	mitochondrial inte
27	268.5	7.3	265	2 G37271	metalloproteinase
28	267.5	7.3	268	2 F37271	metalloproteinase
29	267	7.3	290	2 E37271	metalloproteinase

30	189	5.2	820	2 T27909	hypothetical prote
31	175.5	4.8	615	2 F82132	probable peptidase
32	145.5	4.0	1922	2 T00637	hypothetical prote
33	143	3.9	532	2 C83696	hypothetical prote
34	140	3.8	2166	2 G70163	hypothetical prote
35	139.5	3.8	852	2 D72230	conserved hypothet
36	139.5	3.8	2269	2 T28677	hypothetical prote
37	139	3.8	1084	2 G71329	conserved hypothet
38	137.5	3.7	880	2 F75103	oligopeptidase - H
39	135	3.7	578	2 B71934	polyprotein allerg
40	135	3.7	1095	2 T31423	oligopeptidase
41	133.5	3.6	590	2 H70130	oligopeptidase
42	133	3.6	547	2 H72311	hypothetical prote
43	133	3.6	1115	2 T41342	probable coiled-co
44	133	3.6	1345	2 A64430	DNA-directed RNA p
45	133	3.6	1392	2 A43336	microtubule-vesicl

ALIGNMENTS

RESULT 1  
A43411  
soluble angiotensin-binding protein - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 20-Jun-2000  
C:Accession: A43411  
R:Sugiura, N.; Hagiwara, H.; Hirose, S.  
J. Biol. Chem. 267, 18067-18072, 1992  
A:Title: Molecular cloning of porcine soluble angiotensin-binding protein.  
A:Reference number: A43411; MUID:92388174  
A:Accession: A43411  
A:Molecule type: mRNA  
A:Residues: 1-704 <SUG>  
A:Cross-references: GB:D11336; NID:g217708; PID:BAA01949.1; PID:g217709  
A:Note: sequence extracted from NCBI backbone (NCBIN:112884, NCBIP:112886)  
C:Superfamily: thimet oligopeptidase  
C:Keywords: mitochondrion

Query Match 95.0%; Score 3483; DB 2; Length 704;  
Best Local Similarity 94.5%; Pred. NO. 1.5e-203;  
Matches 665; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

QY	1	MIARCLLAVRSLLRRVGSRIILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR	60
DB	1	MIYRCLSAARLHRVGGSGILLRMTLGRVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR	60
QY	61	TEELIVOTKQYDAVGMGLGIEEVTYENCLQALADVEVKYIVERTMLDFFQHVSSDKQVRA	120
DB	61	TEELIAQTKQYDDIGMLDIEEVTYENCLQALADVEVKYIVERTMLDFFQHVSSDKQVRA	120
QY	121	ASTEADKRLSRFDIEMSRGDIIFERIVHQTCDLGKIKPEARRYLEKSIKMKRNLGHL	180
DB	121	ASTEADKRLSRFDIEMSRREDIFLRIVRLKETCDLGKIKPEARRYLEKSVKMKRNLGHL	180
QY	181	PEQVQNEIKSMKRMSELCLDFNKNLNEEDTFLVFSKAELGALPDDFIDLSLEKTDQDKYK	240
DB	181	PEQVQNEIKAMKRMSELCLDFNKNLNEEDTFLVFSKAELGALPDDFIDLSLEKTDQDKYK	240
QY	241	ITLKYPHYFVPMKKCCIPETRRRWEMAFNTRCKEENTIILOQLPLRTKVKAKLIGYSTHA	300
DB	241	ITLKYPHYFVPMKKCCIPETRRRWEMAFNTRCKEENTIILOQLPLRAKVKALLIGYSTHA	300
QY	301	DFVLEMTAKTSRVTAFDLDLSQKLKPLGEAREFETLNLKKKECKDRGFEGYDGNKINAWD	360
DB	301	DFVLEMTAKTSRHHVTAFLDLDLSQKLKPLGEAREFETLNLKKKECKDRGFEGYDGNKINAWD	360
QY	361	LYIYMTQTEELKYSIDQEFLEKFFPEVTEGLLNTYQELLGLISFEQMTDAHVNNKSVTL	420
DB	361	LHYIYMTQTEELKYSVDQEFLEKFFPEVTEGLLNTYQELLGLISFEQMTDAHVNNKSVTL	420
QY	421	YTVKDKATGEVLQGFYLDLYIPREGKYNHAAACFGLQPCCLLPDGSRMMAVAALVVFNSQPV	480

```

Db 421 YTVKDKATGEVLGQYLDLYPREGKYNHAACFGLOPCLLPDGRSMMSVAALVNFSSQPR 480
QY 481 AGRPSLLRHDEVRTYFHEFGVHMHQICAOQDFARFSGTNNVETDFEVPVSQMLENNWWDVD 540
Db 481 AGRPSLLRHDEVRTYFHEFGVHMHQICAOQDFARFSGTNNVETDFEVPVSQMLENNWWDVD 540
QY 541 SLRLSKHYKDGSPDADLLEKLVASRLVNTGLTLRQIVLSKVDQSLHTNTSLDAASEY 600
Db 541 SLRLSKHYKDGSPDADLLEKLVASRLVNTGLTLRQIVLSKVDQSLHTNTSLDAASEY 600
QY 601 AKYCEILGVAATPCTNNPATFGLHAGYDGYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
Db 601 AKYCEILGVAATPCTNNPATFGLHAGYDGYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
QY 661 VGMKYRNILKPGSGDGMQLHFLKREPNOKAFRLMSRGLHAP 704
Db 661 VGMKYRNILKPGSGDGMQLHFLKREPNOKAFRLMSRGLHAP 704

RESULT 2
A45985
microsomal endopeptidase MEP - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C:Accession: A45985
R:Kawabata, S.; Nakagawa, K.; Muta, T.; Iwanaga, S.; Davie, E.W.
J. Biol. Chem. 268, 12498-12503, 1993
A:Title: Rabbit liver microsomal endopeptidase with substrate specificity for processing
A:Reference number: A45985; MUID:93286083
A:Accession: A45985
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-704 <KAW>
A:Cross-references: GB:D13310; NID:g391848; PIDN:BAA02570.1; PID:g391849
A:Experimental source: liver
A:Note: sequence extracted from NCBI backbone (NCBIN:133580, NCBIIP:133581)
C:Superfamily: thimet oligopeptidase

Query Match 94.7%; Score 3475; DB 2: Length 704;
Best Local Similarity 93.3%; Pred. No. 4.5e-203;
Matches 657; Conservative 30; Mismatches 17; Indels 0; Gaps 0;

QY 1 MIARCLAVRSILRRVGGSRILLRTGLREVMSPLQAMSSYTVAGRNVLRLWDLSPQIKTR 60
Db 1 MIARCFSAVRLHVRGGSRILFKMTLGLREVMSPLQAVSSYTAAGRNVLRLWDLSPQIKTR 60
QY 61 TEELIVQTKQYDVGMLGIBEVTVYENCLQALADVEVYIIVERTMLDPQHVSSDKVRA 120
Db 61 TEELIAQTKQYDVSVMGLDIDKDVYENCLQALADVEVYIIVERTMLDPQHVSTDREYRA 120
QY 121 ASTEADKRLSRFDIEMSRGDIFFERIVHLOETCDLGKTKPEARRYLEKSIKMGKRNGLHL 180
Db 121 ASTEADKRLSRFDIEMSRREDIFORIVHLOETCDLEKTKPEARRYLEKSVKMGRRNGLHL 180
QY 181 PEOVQNEIKSMKKRMSCLIDFNKLNEDDFFLVPSKALGALPDFDIDSLEKTDKKYK 240
Db 181 PEOVQNEIKSMKKRMSCLIDFNKLNEDDFFLVPSKALGALPDFDIDSLEKMDKKYK 240
QY 241 ITLKYPHYFPVMKKCIPETRRRMEAFNTRCKEENTIILOQLPLRTKVAKLIGSYTHA 300
Db 241 ITLKYPHYFPVMKKCIPETRRRMEAFNTRCKEENTVILQQLPLRAQVAKLLIGSYTHA 300
QY 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGENAREFIIINLKKKECKRGFEYDQGINAWD 360
Db 301 DFVLEMTAKTSRVTAFDLDLSQKLKPLGENAREFIIISLKKKECKRGFEYDQGINAWD 360
QY 361 LYXYMTQTEELKYSIDQEFLEKYPFIEVVTEGLLNTYQELLGLSFEQMTDAHVNKSVTL 420
Db 361 LHYMTQTEELKYSIDQEFIEKYPFIEVVTEGLLNTYQELLGLSFEQVADAHVNPSTL 420
QY 421 YTVKDKATGEVLGQYLDLYPREGKYNHAACFGLOPCLLPDGRSMMAVAALVNFSSQPV 480
```

```

Db 421 YTVKDKATGEVLGQYLDLYPREGKYNHAACFGLOPCLLPDGRSMLSVAALVNFSSQPV 480
QY 481 AGRPSLLRHDEVRTYFHEFGVHMHQICAOQDFARFSGTNNVETDFEVPVSQMLENNWWDVD 540
Db 481 AGRPSLLRHDEVRTYFHEFGVHMHQICAOQDFARFSGTNNVETDFEVPVSQMLENNWWDVD 540
QY 541 SLRLSKHYKDGSPDADLLEKLVASRLVNTGLTLRQIVLSKVDQSLHTNTSLDAASEY 600
Db 541 SLRLSKHYKDGSPDADLLEKLVASRLVNTGLTLRQIVLSKVDQSLHTNTSLDAASEY 600
QY 601 AKYCEILGVAATPCTNNPATFGLHAGYDGYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
Db 601 ARYCTDILGVAATPCTNNPATFGLHAGYDGYGYGLWSEVFSMDMFYSCFKKEGINNPE 660
QY 661 VGMKYRNILKPGSGDGMQLHFLKREPNOKAFRLMSRGLHAP 704
Db 661 VGMKYRNILKPGSGDGMQLHFLKREPNOKAFRLMSRGLQAP 704

RESULT 3
S43250
thimet oligopeptidase (EC 3.4.24.15) - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C:Accession: S43250
R:Kato, A.; Sugiyama, N.; Hagiwara, H.; Hirose, S.
Eur. J. Biochem. 221, 159-165, 1994
A:Title: Cloning, amino acid sequence and tissue distribution of porcine thimet oligo
A:Reference number: S43250; MUID:94222057
A:Accession: S43250
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-687 <KAT>
A:Cross-references: EMBL:D21871; NID:g456612; PIDN:BAA04882.1; PID:g565284
C:Superfamily: thimet oligopeptidase
C:Keywords: hydrolase; metalloproteinase; zinc

Query Match 63.5%; Score 2330; DB 2: Length 687;
Best Local Similarity 64.0%; Pred. No. 1.1e-133;
Matches 431; Conservative 104; Mismatches 130; Indels 8; Gaps 1;

QY 29 EVMSPLOAMSSYTVAGRNVLRLWDLSPQIKTRTEELIVQTKQYDVGMLGIEEYVENC 88
Db 13 DVAAPCSAV-----NYLRWDLSSAQOIGELTTEIEQTRVYDRVGTQELQDVSYENT 64
QY 89 LQALADVEVKYIVERTMLDFFQHVSSDKVEVRAASTEADKRLSRFDIEMSRGDIPIRIVH 148
Db 65 LKALADVEVTVQGNILDFFQHVSPCKDIRTASTEADKRLSEFDVEMSRQDVYQRIWV 124
QY 149 LQETCDLGLKIPKPEARRYLEKSIKMGKRNGLHLPQOVQNEIKSMKKRMSCLIDFNKLN 208
Db 125 LQEKVQKDSLREAPARYLERLIKLRNRNGLHLPKETQEKISIKKLSLCLIDFNKLN 184
QY 209 DDTFLVPSKALGALPDFDIDSLEKTDKKYKTKLKYPHYFPVMKKCCIPETRRRMEAF 268
Db 185 DTTFLPVTREELGLPEDFLNSLEKTEDEKLKVTLYKPHYPLPKKKCHVPTRRRVEAF 244
QY 269 NTRCKEENTIILOQLPLRTKVAKLIGYSTHADFFVLENTAKTSRVTAFDLDLSQKLKP 328
Db 245 NCRKEENCAILRELVRRAQKSSLLGFSTHADYVLENNMANKTSQVATFDELAKLKP 304
QY 329 LGAEAREFTILNLKKECKRGFEYDQGINAWDLYYMTQTEELKYSIDQEFLEKYPFIEV 388
Db 305 LGEQERAVILELKAECTKRGDLDFGRINAWDMRYMNOVEETRYRVQNLKKEYFPQV 364
QY 389 VTEGLLNTYQELLGLSFEQMTDAHVNKSVTLTYVKDKATGEVLQYLDLYPREGKYNH 448
Db 365 VTRGLLGYIYQELLGLTFHLEGAAVWHEDVMLYSYRDAASGVKIGFYLDLYPREGKYGH 424
QY 449 AACFGLOPCLLPDGRSMMAVAALVNFSSQPVAGRPSSLRHHDEVRTYFHEFGVHMHQICA 508
```

Db 425 AACFGLQPCLRDGRSQRTAIAAWANFTKPTDAPSLLOHDEVETFFHEFGHVMHQLCS 484

QY 509 QTFARFSGTNGTDEVEVPSQMLENNWVDLSRLSKHYKDGSPDIADLLLEKLVASRL 568

Db 485 QAEFAMFSGTHVERDVEAPSQMLENNWVEAPLLRMSQHYRTGSAIPQELLEKLKRSQ 544

QY 569 VNTGLTLRQIVLSKYDQSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAG 628

Db 545 ANTGLFNLQIVLAKVDQALHTQTAAAPAEYARLCQEILGVPAATPGTNMPATFGHLAG 604

QY 629 YDQYYGYLWSEVFSMDMYSCFKKGINPVGMYRNLLILKPGSLDGMMLHNLFR 688

Db 605 YDQYYGYLWSEVFSMDMYSTRFKQEGILSGKVGMDYRSCILRPGGSEDASVMLKFLGR 664

QY 689 EPNOKAFMLSRGL 701

Db 665 DPQDAFLLSKGL 677

RESULT 4

HYHUTH

N:Alternate names: beta-amyloid precursor protein processing metalloproteinase; metalloproteinase

C:Species: Homo sapiens (man)

C>Date: 27-Sep-1995 #sequence\_revision 17-Nov-1995 #text\_change 18-Jun-1999

C:Accession: J04197; PC4053; G01881; G01882; A53633

R:Thompson, A.; Huber, G.; Malherbe, P.

Biochem. Biophys. Res. Commun. 213, 66-73, 1995

A:Title: Cloning and functional expression of a metalloendopeptidase from human brain with

A:Reference number: J04197; MUID:95367027

A:Accession: J04197

A:Molecule type: mRNA

A:Residues: 1-689 <THO1>

A:Cross-references: GB:Z50115; NID:g1030054; PIDN:CAA90477.1; PID:g1030055

A:Experimental source: brain

A:Accession: PC4053

A:Molecule type: protein

A:Residues: 67-78;181-200 <THO2>

R:Munroe, D.G.

submitted to the EMBL Data Library, June 1995

A:Reference number: H00694

A:Accession: G01881

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-689 <MUN1>

A:Cross-references: EMBL:U29366; NID:g1098599; PIDN:AAA82607.1; PID:g1098600

R:Munroe, D.G.

submitted to the EMBL Data Library, June 1995

A:Reference number: H00695

A:Accession: G01882

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-5 <MUN2>

A:Cross-references: EMBL:U29367; NID:g1098597; PIDN:AAA82606.1; PID:g1098598

R:Papadimitris, G.; Siman, R.; Scott, R.; Abraham, C.R.

Biochemistry 33, 192-199, 1994

A:Title: Identification of a metalloprotease from Alzheimer's disease brain able to degrade

A:Reference number: A53633; MUID:94114484

A:Accession: A53633

A:Molecule type: protein

A:Residues: 67-78;181-197,'X',199-200 <PAP>

A:Experimental source: brain of Alzheimer's disease patient

A>Note: sequence modified after extraction from NCBI backbone

C:Comment: This metalloproteinase is responsible for cleaving a Met-Asp bond in the Alzheimer's

C:Genetics:

A:Gene: GDB:THOP1; GDB:ME78

A:Cross-references: GDB:595011; GDB:642232; OMIM:601117

A:Map position: 19q13.3-19q13.3

C:Function:

A:Description: endopeptidase preferentially hydrolyzing peptide bonds on the carboxyl side

C:Superfamily: thimet oligopeptidase

C:Keywords: Alzheimer's disease; amyloid; brain; glycoprotein; hydrolase; metalloproteinase

F:2-689/Product: thimet oligopeptidase #status predicted <MAT>

F:451/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:473-477,480/Binding site: zinc (His) #status predicted

F:474/Active site: Glu #status predicted

Query Match 63.5%; Score 2330; DB 1; Length 689;

Best Local Similarity 64.9%; Pred. No. 1.1e-133;

Matches 426; Conservative 97; Mismatches 133; Indels 0; Gaps 0;

QY 46 NVLRWDLSPQIKTRTEELIVQTKQYDVGMLGIEBVEYENCLOALADVEVYIVERTM 105

Db 22 NDLRWDLQAQIEERTRELIEQTQKRYDQGTQEFDESVYESTLKALADVEVYTVQRNI 81

QY 106 LDPPQHVSSDKVRAASTADKRLSRFDLEMSMRGDIFFERIVHLQETCDLGKIKPARRY 165

Db 82 LDPPQHVSPSKDITASTADKRLSEDFVEMSMREDVYQRIWVQLQEKVQKDSLRPEARY 141

QY 166 LEKSIRKMGKRNGLHLPEQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKAEALGALPD 225

Db 142 LERLIKLGRRNGLHLPRETQENIKRIKKLLSLCIDFNKNLNEDDTFLPPTLQELGGLPE 201

QY 226 DFIDSLEKTDYDDKYKTLKYPHPVPMKKCCIPETRRMEMAFTNRCKENTIIQLQLLP 285

Db 202 DFLNSLEKMDGKLYTLKYPHPVPLKKCHVPETRRKVEEAFNCRCKEENCAILKELVT 261

QY 286 LRTPKVAKLIGYSTHADPVLMTAKSTSRVTAFLDDLQSLKPLGAEAREFIINLKKKEC 345

Db 262 LRAQKRLILGFTHADYVLEMMNAKTSQVATFDELAQKPLGQERAVILELKAEC 321

QY 346 KDRGFYDGGKINAWDLVYNTQTEELKYSIDQEFLEKEYFP IEVTEGLLNTYQELGLSF 405

Db 322 ERGGLPFGIRAWDMRYNNQVETRYCVDQNLKEYFPVQVTVHGLLGIYQELGLGF 381

QY 406 EQMTDAHVNKSVTLTYTKDQKATGEVLGVLDLYPREGKYNHAAFCGLQPCGLPDGSR 465

Db 382 HHEEGASAMHEDVRLYTARDAASGEVVGKPYLDLYPREGKYGHAAFCGLQPCGLQDGR 441

QY 466 MVAVALVNFSPQVAGRPSLLRHDEVRYTFHEFGHVMHQLCAQTDPAFESGYNVEDFV 525

Db 442 QIAIAAWANFTKPTADAPSLLOHDEVETFFHEFGHVMHQLCSQAEFAMFSGTHVERD 501

QY 526 EVPSQMLENNWVDLSRLSKHYKDGSPDIADLLLEKLVASRLVNTGLTLRQIVLSKYD 585

Db 502 EAPSQMLENNWVQEPFLRMSRYRTGSAVPRELLEKLTIESRQANTGLFNLQIVLAKVD 561

QY 586 QSLHTNTSLDAASEYAKYCEILGVAATPGTNMPATFGHLAGYDQYYGYLWSEVFSMD 645

Db 562 QALHTQTDADPAEYARLCQEILGVPAATPGTNMPATFGHLAGYDQYYGYLWSEVFSMD 621

QY 646 MEYSCFKKGINPVGMYRNLLILKPGSLDGMMLHNLFRKREPQKAFMLSRGL 701

Db 622 MFHTRFQEGVLNKGMDYRSCILRPGGSEDASAMLRFLGRDPKQDAFLLSKGL 677

RESULT 5

HYHUTH

N:Alternate names: endo-oligopeptidase A; endopeptidase 24.15; Pz-peptidase; soluble

C:Species: Rattus norvegicus (Norway rat)

C>Date: 30-Sep-1992 #sequence\_revision 17-Nov-1995 #text\_change 07-May-1999

C:Accession: S38760; A54152; S55999; A36165

R:McKie, N.; Dando, P.M.; Rawlings, N.D.; Barrett, A.J.

Biochem. J. 295, 57-60, 1993

A:Title: Thimet oligopeptidase: similarity to 'soluble angiotensin II-binding protein

A:Reference number: S38760; MUID:94029935

A:Accession: S38760

A:Molecule type: mRNA

A:Residues: 1-687 <CK>

R:Pierotti, A.; Dong, K.W.; Glucksman, M.J.; Orłowski, M.; Roberts, J.L.

Biochemistry 33, 622, 1994

A:Reference number: A54152; MUID:94114549

A:Contents: corrections





```

180 KIKEKKIKSVANSINFSKNLGEQKEYITTKQLEGVDPDSILTQPTIKSDKSDNETLYK 239
241 IYLYKPHYPVPMKCCIPETRRMEMAFNTRCKEENTIILOQLLPLRTKVAKLGYSTHA 300
240 VTFKPIPIFVPMKLASSAQKQAFADQNKV-PENEAILLDTLKLRLDELASLLGYDTYA 298
301 DFVLEMTAKSTSRVTAFTADLDSQKLKPLGEAERFELNKKKECKDRGFEDGKINAWD 360
299 NYNYLDKMAEDSTVWNFLNDLKKLPLGRKELQVLODKMAEDVKLNQGDADPNYYIWD 358
361 LYMYMTQTELYKSIDQELPKYFPEVVEVTEGLLNTYQELGLSFEQMTDAH---VWN-- 415
359 HRYNDKYLLENFNVDLEKISEYFPLEATITGMLEYETFLNLFKTIETKDSQNSVWHD 418
416 -KSVTLTYVKDKATGEVLQGFYLDLYPREGKYNHACFGLOQCGLLPDGRMMVAALVY 474
419 VKOIAVNMDDPKSPNFVGIWYFDLHPRDQKYGHAANFGLSFMIDDTTRSPYATVYC 478
475 NFSQVAGRPSSLRRHDEVRTYFHEFGVNHQICAQTDFAFSGT-NVETDFVEVPSOMLE 533
479 NFSKSTOKPSLLKNEIVTFPELHGCHGHDLVGQNKESRPNPGSPWDFVEAPSOML 538
534 NWVDVDSLRRLSKHYKDGSPITADDLLEKLVASRLVNTGLTLRQIVLSKYDQSLHTNTS 593
539 FWTWKNKELINLSHYKTEKIPESLINSIKTKHVNGALFTLRQLHFGLEDKMKVHTCKD 598
594 LDAAS-EYAKYCSSELLGAATPGTNMPA-TFGH-LAGGYDQYGYLWSEVFSMDMEY 648
599 LQNLISICDTFWNLDRDISLISNGGTSKGYDSFGHIMSDSYAGYGYLWAEVATDMYH 658
649 SCFKKEGINPVEGMYKRNLIILKPGSLDGMHNLHFLKREPQKAFMSRGL 701
659 TFEAKDP-LNANGIOYRDIVLARGGLYDINDNLKFLGREPSKDAFLKELGL 710

RESULT 7
C82353 oligopeptidase A VC0188 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82353
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, B.
  l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
  Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-680 <HET>
A:Cross-references: GB:AE004109; GB:AE003852; NID:g9654590; PIDN:AAF93364.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0188
A:Map position: 1
C:Superfamily: peptidyl-dipeptidase Dcp

Query Match 23.1%; Score 848; DB 2; Length 680;
Best Local Similarity 31.0%; Pred. No. 7.4e-44;
Matches 209; Conservative 128; Mismatches 304; Indels 34; Gaps 12;

QY 52 LSPQIKTTEELIVQTKQYDVGMLGTEEVYENCLOALADVVKYIVERTMLDFPOH 111
DB 17 IKPEHVKAQEAIDCRHTIDKV-LAENPQPSWESVIAPIEVDRLSKWSPVSHMS 75
QY 112 VSSDREVAASFEADKRLSRFDIEMSRGDIERIVHLOETCDLGLKIKPEARYLEKSTK 171
DB 76 VVNSDELREAVESCLPLLSSEYSTWVGQHGKLFYAKTKESAEFAKLDRAQOKNSDSL 135
QY 172 MGKRNGLHLPEQVQVQNEIKSMKRMSELCDIFNKNLNEIDDTFLVFSK-----ELGALPDD 226
```

```

136 DFELSGLGLPQEQRYGEISKRMSSELGSKFSNNVL--DATMGWTQKITDVNLLAGMPES 193
227 FTDLSLEKTDODK----YKITLKYPHYFVPMKCCIPETRRMEMAFNTRC-----KE 274
194 ALAAQAAAAEAAGLEGLYLLTDIPSLPYMTYXCDNALRKEYEAYVYTRASDRGPNAGW 253
275 ENTIILOQLLPLRTKVAKLGYSTHADFLVLEMTAKSTSRVTAFTADLDSQKLKPLGEAER 334
254 DNSEIIAEOLKRLHEIARMLGFSTYSEKSLATKMAQTTDOVLGFLNDLANKAKPQGEREV 313
335 EFILNLKKKECKDRGFEDGKINAWDLYIYMTQTEELKYSIDQELPKYFPEVVEVTEGLL 394
314 EELRQFAESEFGVK-----QLELWDIAIYSEKQKHLFEISDEELRPVPEQKVVGGLF 367
395 NYVOELLGLSFPOMDAHVNKSVLYTYVKDKATGEVLQGFYLDLYPREGKYNHACFG 454
368 EYLSRLFGQVKEKRGQVDMHESVRFDFID-AQGTIRGSGFYLDLYAREHKKRGGAWMDEC 426
455 QPCGLLPDGRMMVAALVWNFSQPVAGRPSSLRRHDEVRTYFHEFGVNHQICAQTDFA 514
427 RYKRTDTSALQTPVAYLTCNPNRPVGDPALETHDEVTLTFHEFGHGHMLTQVEVGA 486
515 FSGTN-VETDFVEVPSOMLENVWDVSLRLSKHYKDGSPITADDLLEKLVASRLVNTGL 573
487 VSGINGVPWDVAVELPSQFLENWCQBEALAFISGHYQTCGEPLPKAMLDKMLAAKNFQ 546
574 LILRQIVLSKYDQSLHTNTSLDAASEYAKYCSSEI-LGVAATPG---TNMPATFGHL-AGG 628
547 FILRQLEGLFDLTLLHTTYDPEVGPVKVLETLAEVKKVAVLPGLGLEWNRFSHSF 606
629 YDQYGYLWSEVFSMDMEYSCFKKEGINPVEGMYKRNLIILKPGSLDGMHNLHFLKR 688
607 YSAGYYSYLWAEVLSADAF-SRFEIEGIFNRETGQSFLLNILEMGGSEEPMEFLFKRPR 665
689 EPNQKAFMSRGLHA 703
666 EPQIDALLRHAGIAA 680

RESULT 8
S47718
Oligopeptidase A (EC 3.4.24.-) - Escherichia coli
N:Alternate names: prIC protein
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 29-Sep-1999
C:Accession: S47718; E65147; A43329
R:Plunkett, G.
  submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47718
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-680 <PLD>
A:Cross-references: EMBL:U00039; NID:g46582; PIDN:AAB18474.1; PID:g466635
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
  Riley, M.;
  Rose, D.J.; Mau, B.; Shao, Y.
  Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65147
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-680 <BLAT>
A:Cross-references: GB:AE000426; GB:U00096; NID:g1789910; PIDN:AAC76523.1; PID:g17899
A:Experimental source: strain K-12, substrain MG1655
R:Conlin, C.A.; Trun, N.J.; Silhavy, T.J.; Miller, C.G.
  J. Bacteriol. 174, 5881-5887, 1992
A:Title: Escherichia coli prIC encodes an endopeptidase and is homologous to the Salm
A:Reference number: A43329; MUID:92394891
A:Accession: A43329
A:Status: preliminary
A:Molecule type: DNA
```

A;Residues: 1-42, 'E', 44-210, 'FW', 213-264, 'AA', 267-389, 'H', 391-405, 'T', 407-515, 'G', 517-688  
A;Note: sequence extracted from NCBI backbone (NCBIN:113030, NCBIIP:113031)  
C;Genetics:  
A;Gene: prlc  
C;Superfamily: peptidyl-dipeptidase Dcp  
C;Keywords: hydrolase; metalloproteinase

Query Match:	22.3%;	Score 818.5;	DB 2;	Length 680;
Best Local Similarity	31.9%;	Pred. No. 4.6e-42;		
Matches 206;	Conservative 122;	Mismatches 276;	Indels 41;	Gaps 15;

[illegible]

RESULT 9 |  
 B86021  
 oligopeptidase A [imported] - Escherichia coli (strain O157:H7)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
 C:Accession: B86021  
 R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 iller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference: number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: B86021  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues:11-680 <STO>  
 A:Cross-references: GB:AE0005174; NID:q12518192; PIDN:AAG58630.1; GSPDB:GN00145; UWGP:248

A;Experimental source: strain 0157:H7, substrain EDL933  
C;Genetics:  
A;Gene: prfC

Query Match 22.3%; Score 817.5; DB 2; Length 680;  
Best Local Similarity 31.9%; Pred. No. 5.2e-42;  
Matches 206; Conservative 122; Mismatches 276; Indels 41

Qy	84	TYENCLOALADV	VEKVIYIVERT	MDLFPQH	VSSDK---	EVRAASTEAD	KRLSRFDIEM	SMRG	140
Db	48	TWENLCQPLAE	VD---DVLGR	IFSPVSHL	SVNKS	PELREAY	EQTLPLLSY	STWV	104
Qy	141	DIFERIVHLO	ETCDLKIR	PEARRY	LEKSIKMK	RNGLHL	PEQVONEI	KSMKRN	200
Db	105	GLYKAYRDL	RHDGHYAT	LNTAQAKA	VDNALR	DELSSIG	LPGKEQORY	GEIATRL	164
Qy	201	DFNKNLNE	DD---TFLV	FSKAEI	ALPD	DFID	SLKRTDD	K---YK	253
Db	165	QYSNNVL	DATMGW	TKLVT	DEALAG	MPESAL	AAAKAA	EKALEGY	224
Qy	254	KCCIPETRR	REMAFN	TRC-----	KENT	IILOQL	PLRTK	VAKILGY	305
Db	225	YCDNQVL	REEMY	RAYSTR	ASDQ	GNAG	KWDNS	MMEEI	284
Qy	306	MNTAKTS	SRVTA	FLDDL	SOKL	KPLG	BAEREF	ILNLK	365
Db	285	TKMAEN	QOVL	DFEL	TLAKR	AR	PQGK	ELAQL	338
Qy	366	TQTEELK	YSID	QELF	KYPIEV	VTGL	NTYQ	ELLGL	425
Db	339	EKOKHLY	SISD	QLRPY	FPEN	KA	VNGLF	EVVKRY	398
Qy	426	KATGEVL	GOY	LDLY	PREG	YNH---	ACFGL	QPGCL	482
Db	399	E-NNELG	RSYD	ILY	AREN	KRGW	MMDC	QVMRKA	454
Qy	483	RPSELLR	HDEV	RTYTF	HEFB	GHMHO	ICAQ	TDFAR	541
Db	455	KPALF	THDEV	ITL	FEHFG	HGLH	MLTRI	ETAG	514
Qy	542	LRRLSKH	KDGS	PIADD	LLLEK	LVAS	RLVNT	GLTL	601
Db	515	LAPITSGH	YETGE	PLPKEL	DLKML	AAANY	QAALF	ILRQLE	574
Qy	602	KYCS	EILG-VAAP	PGTN---	MPAT	FCHL-AGG	YDGOY	YGLW	656
Db	575	ETLAEIK	LVA	VVPS	PSWG	FFPHAF	SHI	FAGG	633
Qy	657	MNPEVGM	KMYRN	LILK	PGSLDG	MDML	HNFL	KLRPN	701
Db	634	FNRETQ	SF	DLN	ILSRG	SGSE	EPMD	LFR	678

RESULT 10

C64055  
thimet oligopeptidase (EC 3.4.24.15) - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae

C:\Program Files\Microsoft Office\Office\Word\Word.doc

#sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999

C/Accession: C64055  
C/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
R.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
D.M.; Brandon, R.C.; Fines, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995

A:Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

A: Reference number: A64000; MUID: 95350630

A;Accession: C64055

A; Status: nucleic acid

A;Molecule type: DNA

A;Residues: I-68I <TIGR>

A;CROSS-References: GB:032706; GB:L42023; NID:91373170; FIDN:AAC21002.1; FID:91373175  
C:Superfamily: peptidyl-dipeptidase Dcp

C:Keywords: hydrolase; metalloproteinase

Query Match 22.2%; Score 814.5; DB 2; Length 681;  
Best Local Similarity 30.3%; Pred. No. 8e-42; Mismatches 127; Matches 206; Conservative 127; Indels 49; Gaps 15;

QY 52 LSPQIKTRTEELIVQTKQVDAVGMGLGTEEVYENCALQALADVEVKYIVERTMLDFPOH 111  
DB 19 IKPEHRAVEKLQDCRNTIQV--LKPHETWENFILPTETNDRNLRAWSPVSHLNS 76

QY 112 VSSDEVRRASTEADKRSLRFDIEMSRGDIPIERIVHLOETCDLGKIKPEARYLEKSTK 171  
DB 77 VKNSTELREAYOTCLPLLSEYSTWVGQHGKLYNAYLAKNSAEFADYSTAQKKAATENSUR 136

QY 172 MGRKRLHPEQVQNEIKSMKRMSELCIDFNKNLNEEDTF---LVPSKAEGLGALPDF 227  
DB 137 DFELSGIGLSEKQQRGYGEIVARLSNLSQFSNNV-LDAMGWELKIENAEALAGLPESA 195

QY 228 IDSLEKTDODK----YKILKYPHYFPMKKKCCIPETRRRMEAFNTRCKE-----E 275  
DB 196 LQAQQAFAKSGKLGKGYRFTLEIPSLPVMTYCNALREEMVYRAYATRASEOGPNAGKWD 255

QY 276 NYIILOQLPLTKVAKLGYSTHADVLEMTAKTSRVTAFLDDLQKLPGEARE 335  
DB 256 NSGMVEEILTLELAKLIGFNTYELSLATKMAENPQOVLDLHLAERAKPQGEKELQ 315

QY 336 FTLNKKKECKDRGEYDQ--KINAWDLYYMTQTEELKYSIDQEFLEKYPFIEVVTEGL 393  
DB 316 ELKGYCEKEF-----GVTELAPWDIGFYSEKQKQHLIYAINDBELRPYPFENRVISGL 367

QY 394 LNTYQELGLSFEQMTDAHVNMKSVTLTYVKDKATGEVLQGYLDLYPREGYNHA---A 450  
DB 368 FELIKRIENIRAVEKGVDTWKHVRFPDLIDE-NDQLRGSEYLDLYAREHKGAWMD 426

QY 451 CRLGPGCLLPDGRMMAVAALVNFSPQVAGRPSLLRHDEVRTYFHEFGHYMHQICAO 510  
DB 427 CJCRRKRL---DGSIEPTVAYLTCNFPNAPIGNKPALEFTHNEVTTLTFHEFGHHLMTQ 483

QY 511 DPARFSGT-VETDFVEVPSQMLENNWVDVSLRLSKHYKDGSPDADLLKLVASRLV 569  
DB 484 DVSVDVAGINGVPWDVAVELPSQFMENWCEEAALAFISGHYETGEPLPKELTQLLAKNF 543

QY 570 NTGLTLRQIVLSKYDQSLH-----NTSLDAASEYAKYCEILGVAATPGTNPATF 622  
DB 544 QAAMFIRLQLEFGIFDRLHHTFDKNTQILDLSKVSQVAVIKGV--DWARAPHSF 600

QY 623 GHL-AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNRLILKPGGSLDGM 681  
DB 601 SHIFAGGYAAGYSYLWAEVLSADA-YSRFEEGIFNPTIGKSFSLDEILTRGGSSEPMEL 659

QY 682 LHNELKREPKNQAFILMSRGL 701  
DB 660 FKRRGREGPQLDALLRHKGI 679

RESULT 11  
A42298  
thimet oligopeptidase (EC 3.4.24.15) - Salmonella typhimurium  
N:Alternate names: endo-oligopeptidase A  
C:Species: Salmonella typhimurium  
C:Date: 10-Jul-1992 #sequence\_revision 10-Jul-1992 #text\_change 29-Sep-1999  
C:Accession: A42298  
R:Conlin, C.A.; Miller, C.G.  
J. Bacteriol. 174, 1631-1640, 1992  
A:Title: Cloning and nucleotide sequence of opdA, the gene encoding oligopeptidase A in  
A:Reference number: A42298; MUID:92165739  
A:Accession: A42298  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-680 <CON>  
C:Superfamily: peptidyl-dipeptidase Dcp

C:Keywords: hydrolase; metalloproteinase

Query Match 22.0%; Score 806; DB 2; Length 680;  
Best Local Similarity 30.4%; Pred. No. 2.6e-41; Mismatches 129; Matches 206; Conservative 129; Indels 42; Gaps 17;

QY 52 LSPQIKTRTEELIVQTKQVDAVGMGLGTEEVYENCALQALADVEVKYIVERTMLDFPOH 111  
DB 17 IKPEHVAVTKALADCAAEVGVVHAG-APYSWENLCOPLA--EADDVLGRIFSPI-SH 72

QY 112 VSSDK---EVRAASTEADKRSLRFDIEMSRGDIPIERIVHLOETCDLGKIKPEARYLEK 168  
DB 73 LNSVKNSELPREAYEQTLPLLSEYSTWVGQHGKLYNAYRLDGDYHATLNTAQKKAVDN 132

QY 169 SIKMKGRKRLHPEQVQNEIKSMKRMSELCIDFNKNLNEED---TFLVFSKAEGLGALPD 225  
DB 133 AUURFELSGIGLPEKQQRGYGEIATRLSELGNQYNNVLDATMGWTKLITDEAEELAGMPE 192

QY 226 DFIDS-----LEKTDODKYYKITLKYPHYFPMKKKCCIPETRRRMEAFNTRC-----K 273  
DB 193 SALAAKAAQAEAKQEGYLLITLIDPSYLPVMTYCDNOALREEMVYRAYSTRASDOGNACK 252

QY 274 EMTIILQQLPLTKVAKLGYSTHADVLEMTAKTSRVTAFLDDLQKLPGEARE 333  
DB 253 WNSPVMEEILTLELAKLIGFNTYELSLATKMAENPQOVLDLHLAERAKPQGEKELQ 312

QY 334 REFILNKKKECKDRGEYDQ--KINAWDLYYMTQTEELKYSIDQEFLEKYPFIEVVTEGL 393  
DB 313 LAQLRAFAKEF---GVE---ELQPMDIAYYSEKQKQHLYSISDEQLRPYFENKAVNGL 366

QY 394 LNTYQELGLSFEQMTDAHVNMKSVTLTYVKDKATGEVLQGYLDLYPREGYNHA---A 450  
DB 367 FVAVRIYITAKERTDVDVHPEVRFPELYDE-NNELRGSEYLDLYAREHKGAWMD 425

QY 451 CRLGPGCLLPDGRMMAVAALVNFSPQVAGRPSLLRHDEVRTYFHEFGHYMHQICAO 510  
DB 426 CVQGMKRA---DGLQKPVAYLTCNFPNVPNGKPALEFTHNEVTTLTFHEFGHHLMTRI 482

QY 511 DPARFSG-TNVTDFVEVPSQMLENNWVDVSLRLSKHYKDGSPDADLLKLVASRLV 569  
DB 483 ETAGVSGISGVPMWDVAVELPSQFMENWCEEAALAFISGHYETGEPLPKELDKMLAANY 542

QY 570 NTGLTLRQIVLSKYDQSLHTNTSLDAASEYAKYCEI-LGVAATPGT---NMPATFGHL 625  
DB 543 QAALFILQLEFGIFDRLHAEFNPQOGAKILETLEIKKQVAVVPSPTWGRFPAFHSI 602

QY 626 -AGGYDQGYGYLWSEVFSMDMFYSCFKKEGIMNPEVGMKYNRLILKPGGSLDGM 684  
DB 603 FAGGYAAGYSYLWADVLAADA-YSRFEEGIFNRETGQSFLLDNLITRGGSEPEMELFKR 661

QY 685 FLKREPKNQAFILMSRGL 701  
DB 662 FRGREGPQLDAMLEHYGI 678

RESULT 12  
E83636  
oligopeptidase A PA0067 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83636  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83636  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-681 <STO>  
A:Cross-references: GB:AE004446; GB:AE004091; NID:g9945886; PIDN:AGC03457.1; GSPDB:GN

A;Experimental source: strain PA01  
C;Gene: prlC; PA0067  
C;Superfamily: peptidyl-dipeptidase Dcp

Query Match 21.7%; Score 794.5; DB 2; Length 681;  
Best Local Similarity 30.1%; Pred. No. 1.3e-40;  
Matches 202; Conservative 132; Mismatches 305; Indels 31; Gaps 12;

```
Qy 52 LSPQIKTRTEELIVQKQYDVGMLGIEBVTYENCLOALADVEKVIYVTRTMDLDPFH 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 IRPEHVPAIERILADNRRAAIAARLETRQETQWKLVLAMDELNDRLGAWSVSHLNA 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 112 VSSDKVEYRAASTADKRLSRDIEMSRGDIIFERIVHLQETCDLQKIKPEARRYLEKSIK 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 VCNSEALREAYEACLPESAYSTELGQNRLAFEAYEALAKSPAEGFDVAQKTHLEHALR 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 172 MGKRNGLHLPQVQNEIKSMKKRMSELCDENKML---NEDDTFLVFSKAEGLALPDDFI 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 DFLSGIDLPADKQKRYAEQVSRSELGSRFSNQLLDATQATKHTVTDAAALAGLTSK 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 229 DSEKTD---DKYKTLKYPHPVPMKKCCIPETRRRMEMAFNTRCKE-----EN 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 AQMKAAEAKGLDGLWLSLEPSPSYAVYATYADRALREEVYAACTRASDQGNAGQND 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 277 TIILQQLPLRTKAKLLGYSTHADFVLEMTAKSTSRVTAFLDQLSOKLPLGEAREF 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 GPVNEEILDLQELAGLLGFANYAELSLATKMAESSQVLSFLDLAVRSKPPFAARDLEQ 317
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 337 ILNLKKECKDRGFYDQKINAWDLYYMTQTEELKYSIDQEFLEKYPPIEVVTEGLNT 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 318 L---RAYAAEQGCT---ELQSWDAGYVYAEKLEARYSVSQEALRAVFPVSKVLSGLFAI 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 397 YQELLGSFEQMTDAHVNNKSVITLYTKDKATGVLGGFYLDLPREGKYNHACFGLOP 456
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 VERLYGTOIRLDFHRWADRVLFETLE---NGEHRGFYFDLYARANKRGGAWMGDARD 428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 457 GCLLPDGRMMAVALVNFQVAGRPSSLRHDEVRTYFHEFGHVMHOICAQDFARFS 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 429 RRRDAQGLIDPVAYLCNFTTPVANGRPALLTHDVTFLFHEFGHGLHLTLRVHEHAAS 488
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 517 GTN-VETDFVEVPQOMLNWVDLSRLSKHYKDGSPIDADDLLEKLVALSRLVNTGLLT 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 489 GINGVAMDVAVELPQFMENWCWEPEGLALISAHYETGVVALPQDLLEKMLAAKNFQSGMM 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 576 LRQIVLSKVDQSLTNTSLDRAA---EYAKYCEILGVAAATPCTNNPA-TFGHL-AGGYDG 631
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 VRQLEFLDFELHA--THGDGRSVLQVLEGRIRDEVAVMRPPAYNRFANSFAHIFAGYAA 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 632 QYGYLWSEVSMDFSCFKKEGIMNPEVGMKYNRLILKPGSLDGMDLNHLNFKREPN 691
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 GYYSYKNAEVLSDADF-SRFEEGVFNPDGTGRAFEALARGSGREPMLLPVDFRGREPS 666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 692 QKAFMLSRGL 701
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 667 IDALLRHSGL 676
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 13  
B81224  
oligopeptidase A NMB0214 [imported] - Neisseria meningitidis (strain MC58 serogroup B)  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: B81224

R;Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Hart, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A;Reference number: B81000; MUID:20175755  
A;Accession: B81224

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-678 <TET>  
A;Cross-references: GB:AE002379; GB:AE002098; NID:g7225435; PIDN:AAF40670.1; PID:g722  
A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB0214  
C;Superfamily: peptidyl-dipeptidase Dcp

```
Query Match 19.1%; Score 702; DB 2; Length 678;  
Best Local Similarity 28.4%; Pred. No. 5.2e-35;  
Matches 194; Conservative 126; Mismatches 302; Indels 60; Gaps 19;  
Qy 55 EQIKTRTEELIVQY--KQVYDAVGMGLTEEVY-YENCLOALADVEKVIYVTRT-----ML 106  
Db 16 DOIKTEDIKPAQTAIAEAREQIAAIAKTAQTGTGANTVEPLTG-----ITERVGRINGVV 70  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 107 DFPQHVSSDKVEYRAASTADKRLSRDIEMSRGDIIFERIVHLQETCDLQKIKPEARRYL 166  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 71 SHLSAVADTPELRAYVNELMPEITVFFTEIGQDIELYNREKTIKNSPEFDTLSPAQKTKL 130  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 167 EKSIMKRNGLHLPQVQNEIKSMKKRMSELCDENKML---NEDDTFLVFSKAEGLAL 223  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 131 NHDRLDFVLSGAELVPEQQAELAKLQTEGAOLSAKFSQNVLDATDAFGIYFDDAALAGI 190  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 224 PDD---FIDSLEKTDQDDKYKITLKYPHYPPVPMKKCCIPETRRRMEMAFNTRCKE----- 274  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 191 PEDALAMFAAAQSESKYGYKIGLQIPIYLAVIQVADNRELREQIYRAYVTRASELSDDG 250  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 275 --ENTIILQQLPLRTKAKLLGYSTHADFVLEMTAKSTSRVTAFLDQLSOKLPLGEA 332  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 251 KFDNTANIDRTLNALQTKLLGFKNYAELSLATKMAQDTPEQVLNFLHDLARRAKPYAEK 310  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 333 E-----REFIINLKKCKDRGFYDQKINAWDLYYMTQTEELKYSIDQEFLEKYPF 384  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 311 DLAEVKAFARE-SUNL-----ADLPQWDLGYASEKLEAKYAFSETEVKKYF 356  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 385 PIEVVTGLNTYQELLGSLSPQMTDAHVNNKSVITLYTKDKATGVLGGFYLDLPREG 444  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 357 PVGKVLNGLFAQIKKLYGIGTEKT-VPVHKDVRYFELQ--NGETIGGYMDLYAREG 413  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 445 KYNHAAQFGLOPQCLLPDGRMMAVALVNFQVAGRPSSLRHDEVRTYFHEFGHVMH 504  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 414 KRGAWNDYKGRRRFSDGTQLQPLTAYLVCFAPVPVGGREARLSHDEILFLFHEFHGHLH 473  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 505 QICAQTDFARESGTN-VETDFVEVPQOMLNWVDLSRLSKHYKDGSPIDADDLLEKL 563  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 474 HLLTQVDELGVSGINGVWDVAVELPQFMENWCWEPEGLALISAHYETGVVALPQDLLEK 533  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 564 VASRLVNTGLTLRQIVLSKVDQSLTNTSLDRAA---EYAKYCEI---LGVAATPCTNNPA 620  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 534 LAKNFQGMFLVRQMEFALFDMMIYSEDDSGRLKNMQOVLDSVRKKVAVIQPPEYNRFA 593  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 621 -TFGHL-AGGYDGOYGYLWSEVSMDFSCFKKEGIMNPEVGMKYNRLILKPGGSLDG 678  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 594 LSFHGIFAGGYAGYYSYAWAEVLSDADA-YAAFEESDDV-AATGKRFWEQLIYAVGGSRSA 651  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Qy 679 MDMLNHLNFKREPNOKAFMLSRG 700  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :  
Db 652 AESPKAFRGREPSIDALLRHSG 673  
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 14  
G81996  
oligopeptidase A (EC 3.4.24.70) NMA0054 [imported] - Neisseria meningitidis (strain Z  
C;Species: Neisseria meningitidis  
C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C;Accession: G81996  
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491



us-09-833-782-2.rpr

Tue Jan 15 09:01:18 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:54:36 ; Search time 18.45 Seconds  
(without alignments)  
1371.775 Million cell updates/sec

Title: US-09-833-782-2  
Perfect score: 3668  
Sequence: 1 MIARCLLAVSLRRVGGSR.....FLKREPNOKAFILMSRGLHAP 704

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 170028 seqs, 35950645 residues

Total number of hits satisfying chosen parameters: 170028

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*  
1: /cgn2.6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2.6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2.6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2.6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2.6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2.6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2.6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	21.7	680	5	US-09-897-516-6500
2	732	20.0	655	5	US-09-897-516-6498
3	702	19.1	678	5	US-09-806-866-53
4	702	19.1	678	5	US-09-806-866-57
5	694	18.9	678	5	US-09-806-866-55
6	694	18.9	678	5	US-09-806-866-61
7	641.5	17.5	678	6	US-10-015-127-11537
8	550	15.0	691	5	US-09-897-516-8169
9	480.5	13.1	476	5	US-09-806-866-51
10	475	12.9	491	5	US-09-806-866-59
11	373	10.2	71	5	US-09-826-734-44
12	338.5	9.2	251	6	US-10-015-127-11338
13	130	3.5	1144	5	US-09-708-427-15046
14	130	3.5	1304	5	US-09-708-427-15045
15	130	3.5	1313	5	US-09-708-427-15044
16	122	3.3	1388	5	US-09-976-594-296
17	120	3.3	464	6	US-10-017-754-1906
18	117	3.2	1014	5	US-09-708-427-19883
19	117	3.2	1018	5	US-09-708-427-19882
20	117	3.2	1269	5	US-09-708-427-19881
21	116.5	3.2	701	5	US-09-708-427-3743
22	116.5	3.2	772	5	US-09-708-427-3743
23	115	3.1	1270	5	US-09-720-934-107
24	112.5	3.1	865	5	US-09-815-242-11536
25	111	3.0	961	5	US-09-708-427-12954
26	111	3.0	1027	5	US-09-708-427-12953

ALIGNMENTS

RESULT 1

US-09-897-516-6500  
; Sequence 6500, Application US/09897516  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Huesing, Joseph E.  
; APPLICANT: Krasomil-Osterfeld, Karina C.  
; APPLICANT: Malvar, Thomas M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Spiridonov, Sergei  
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
; FILE REFERENCE: 38-21(51847)B  
; CURRENT APPLICATION NUMBER: US/09/897,516  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215, 161  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 8409  
; SEQ ID NO 6500  
; LENGTH: 680  
; TYPE: PRT  
; ORGANISM: Xenorhabdus sp.  
US-09-897-516-6500

Query Match 21.7%; Score 795; DB 5; Length 680;  
Best Local Similarity 30.2%; Pred. No. 9e-47;  
Matches 207; Conservative 126; Mismatches 295; Indels 58; Gaps 17;

QY 51 DLSPQIKTRTEELIVQTKOVYDAVGMGLTEETVYCNCLALADVEVKYIVERTMLDFQ 110  
DB 16 DIRPWHVPAVKETLANRYQVIEKI-LTENSFTWSLCOPLSEQDKLSRMSPVGHNL 74  
QY 111 HWSSDKVEAASTEADKRLSRFDIENSMRGDIFERIVHQLQETCDLQKPKPEARYLEKSI 170  
DB 75 SVKNSPELREAYEQSLPLLESEFTWLGQHKGLYQAYKSESEVFNLSLQPKRSIEDTL 134  
QY 171 KMKRNGHLHPEQVQNEIKSMKRMSELCLDFNKLNEDDTF----LVFSKALGALPDD 226  
DB 135 RDELSGIGLPEEKQRYKRGYGEIAARMSELGSGQGNV-LDATMGWSKLITDENDLAGLPS 193  
QY 227 ----FIDSLEKTDQDDKVKITLYPHYFPVYMKKCCIPETRRMEMAENTRC-----KE 274  
DB 194 AKAAKAAAEKQEGWLLFLTNMPSYLPVNAVADNRELROEMTHAYTTRASDQGPAGKW 253  
QY 275 ENTIIQLQLPLTKVAKLIGYSTHADFVLEMTAKTSRVTAFLLDLSQKLPAGEAR 334  
DB 254 DNNAIAMEALRHLEAQLLGFASYAEKSLATKMAKNPQQVLDLNDLANRAHGQCKKEL 313

Sequence 449, App  
Sequence 12952, A  
Sequence 512, App  
Sequence 11364, A  
Sequence 19314, A  
Sequence 19313, A  
Sequence 15312, A  
Sequence 515, App  
Sequence 11489, A  
Sequence 9205, Ap  
Sequence 9204, Ap  
Sequence 9203, Ap  
Sequence 21161, A  
Sequence 21160, A  
Sequence 21159, A  
Sequence 4897, Ap  
Sequence 2853, Ap  
Sequence 2854, Ap  
Sequence 2853, Ap

```
QY 335 EFILNKKKECKDRGFYDGG--KINAWDLYYMTOTTELKYSIDQEFLEKXYPPIEVVTEG 392
Db 314 E-----ELTDFARSHVGVKDLKAWDTYYSEKQKHHSIDDEQLRPYPPEORAVEG 365
QY 393 LUNTYQELGLGSFQMTDAHVNKSVTLTYVKDKATGEVLGYQLDLYPRECKYNHA---449
Db 366 LFEVVRRIYGITAKERNVDVETWHSVDVRFELYDE-KGELRGSYLDLYAREHKRGAWMD 424
QY 450 ACFGLOPCGLLDGSRMAVAALVNFQVPAGRPSLLRHDEVRTYTFHEFGHVMHQICAQ 509
Db 425 DCAGRM---REFASGELQKPVAYLTNCFNKPVGDKPALFTHDEVITLTFHEFGHGLHMLTE 481
QY 510 TDFARESGTN-VETDFVEVPSQMLENWWVDVSLRLSKHYKDGSPADDDLEKLVASRL 568
Db 482 IETLDVAGINGVPWDVAVQSPQFMENWCEPALAFISGHVETNEPLQPEMDLNNLAARN 541
QY 569 VNTGLTLTRQIVLSVDQSLTHTNTSLDAASEY--AK-----YCSEILGVAATPGTN 617
Db 542 YQSAFILLRQLEFGLDFRLH-----AEYDPAKGAQILPTLYSVKEQVSVVSPAWN 593
QY 618 -MPATFGHL-AGGYDQGYGYLWSVFSMDYSCFKKEGIMNPEVGMKYNRLILKPGGS 675
Db 594 REPNSFSHFNGGYAAGYSYLMADVLAADS-YSRFSEEGIFNRTTQGSFLDNILSRGGS 652
QY 676 LDGMDLHNLKREPKNQKAFMSRGL 701
Db 653 EDPWTLFTRFRGKPLDMLKSAGI 678

RESULT 2
US-09-897-516-6498
; Sequence 6498, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6498

Query Match 20.0%; Score 732; DB 5; Length 655;
Best Local Similarity 29.3%; Pred. No. 1.8e-42;
Matches 195; Conservative 123; Mismatches 308; Indels 40; Gaps 17;

QY 62 BELIVQTKQYDAVGMGLGIEVYENCIALADVEVKYIVERTMLDFPQHVSDDK---EV 118
Db 2 KEILAKYQVTEKV-VNENSQFTMDNLCPISSEARNQF--ERVWSPV-SHLSVKNTPEF 57
QY 119 RAASTEADKRLSRFDIENSMRGDIFERIVHLOETCDLGIKIPPEARLYLEKSKWKGRNGL 178
Db 58 RVAYEQCLPMNSEFTWNGQKHGLYQAYKSLKESFEFEKLSQPKAVEDKLLDFTLSGI 117
QY 179 HLPQOVQNEIKSMKRMSELCTDKNKNLNEDD---TFVFSKAEGLALPDPDFIDSL- 232
Db 118 ALPEKQRYGEITARLSLSARFGNNVLDATMGWTKLVTDVCDLSGLPENIKSSAKAAA 177
QY 233 KTDDBK-YKITLKYPHYFPVMMKKCCIPETRRMEMAFNTR-----CKEENTIILOOL 283
```

```
Db 178 KAKDLKGLMLTLDPPVYSAVMSYADDCCELREEVYTYAWTRASDOGMNACKWDNNPIMEEI 237
QY 284 LPLRTKVAKLGLGYSTHADFPVLEMTAKSTSRVTAFLDLDLSQKLKPLGAEAREFINLAKK 343
Db 238 LALRHELALLGLGKYNASLAKMTTKKEYLDPLFNLGLVDRAHGQSGREMEELCEFAQS 297
QY 344 E-CKDRGFYDCKINAWDLYYMTOTTELKYSIDQEFLEKXYPPIEVVTEGLLNTYOELLG 402
Db 298 HCYVD-----ELESWDVRYSEKQKHHSIDDEQLRPYPPEORAVEG 350
QY 403 LSFQETMDAHVNKSVTLTYV-KDKATGEVLGYQLDLYPRECKYNHAACFGLQPCGLLP 461
Db 351 ITAKERNVDVETWHSVDVRFELYDE--ELLSGFYMDLYARAGKNGAWQSSVYDMRLA 408
QY 462 DGSRMMAVAALVNFQVPAGRPSLLRHDEVRTYTFHEFGHVMHQICAQDFAFSGTN-V 520
Db 409 SGELQKPVAFINCFNPSGPAKKPALFTHSEYSTLTFHEFGHGLQTLTTIETLDVAGTNGV 468
QY 521 ETDFVEVPSQMLENWWVDVSLRLSKHYKDGSPADDDLEKLVASRLVNTGLTLTRQIV 580
Db 469 PWDVTECSQIMENWCWEABEALEFISGHYETNAPLPKEMLDNMLEAKNYQAAMKMLRQLE 528
QY 581 LSKVDQSLTHTNTSLDAASEYAKYCYSEIL-GVAATPGTN--MPATFGHL-AGGYDQGYG 635
Db 529 FGLDFRLHVEYDPEGAQILPILSVKEHVSVPSSDWERFPNSFSHFNGGYAAGYS 588
QY 636 YLWSEVFSMDYSCFKKEGIMNPEVGMKYNRLILKPGSLDGMMDLHNLKREPKNQKAF 695
Db 589 YMWADVLAADA-YSRFSEEGIFNRTTQGSFLDDYLGRRGSEDPLTLFTRFRGKPLDAL 647
QY 696 LMSRGL 701
Db 648 LKRLGI 653

RESULT 3
US-09-806-866-53
; Sequence 53, Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tetelin, Hervé
; APPLICANT: Venter, J. Craig
; APPLICANT: Maignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Rappuoli, Rino
; APPLICANT: Piazza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-806-866-53
```



```
Query Match          19.1%; Score 702; DB 5; Length 678;
Best Local Similarity 28.7%; Pred. No. 2.2e-40;
Matches 194; Conservative 127; Mismatches 305; Indels 50; Gaps 18;

QY 56 QIKTRTEELIVQTKQYVDVAVGMLGIEEVT-----YENCLOALADVEVKYIVERT-----ML 106
DB 17 QIQTEDIKPAVQT-AIAEARGQIAAQAQTHGTGWANTVEPLTG-----ITERVGRWGVV 70
QY 107 DFPQHVSSDKVEVRAASTADKRLSRFDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYL 166
DB 71 SHLNSVVDPELRVAVNELMPEITVFTFTEIGQDIELYNRFKTIKNSPEFATLSPAQKTLL 130
QY 167 EKSIRMGKRNGLHLPEOVQNEIKSMKRSELCIDFNKN-LNEDDTFLVF--SKAELGAL 223
DB 131 DHDLRDVLVSGAELPERQAEALQTEGAQLSAKFSQNVLDATDAFGYFDDAAPLAGI 190
QY 224 PDD-----FIDSLEKTDQDKYKITLKYPHYFPVWKKCCIPETRRMEMAFNTRCKE----- 274
DB 191 PEDALAMFAAAQSEKTKYKIGLQIPHYLAVIQVAGNRELREQIYRAVTVTRASELSNDG 250
QY 275 --ENTILQOLLPLRTKVAKLGLGYSTHADFLVEMNTAKSTSRVTAFLDLSOKLPLGEA 332
DB 251 KFDNTANIDRTLLENALKTAKLLGFKNYAELSLATKMDTPEQVLFNHLDRARRAKPYAEK 310
QY 333 EREFILNLKKKCKDRGFYDQKIN--AMDLYVYMTQTEELKYSIDQEFLEKVPFLEVY 390
DB 311 D-----LAEVKAFARE-SLNL-----ADLPQWDLGVASEKLEAKYAFSETEVKYKF 356
QY 391 EGLLNTYQELLGLSFQMTDAHWNKSVTLTYVKDKATGEVLGQFVLDLYPREGYNHAA 450
DB 363 AGLFAQIKKLYGIGFAEKT--VPVWHKDVRYFELQ--NGKTIGGVYMDLYAREGRGGAW 419
QY 451 CRGLQPCLLPGSRMMAVAALVNFPSQVAGRPSLLRHDEVRTTFHEFGHVMHQAICAT 510
DB 420 MNDYKGRRRFADGTQLPTAYLVNCFNFPVPGVKGEARLSHDEITLTFHETGHGHLHLLTV 479
QY 511 DFRFSGTN-VETDFVEVPSQMLENNWVDVSLRLSKHYKDGSPITADDLLEKLVASRLV 569
DB 480 DELGVSGINGVENDAVELPSQMFNFVWYNVLAQMSAHEETGEPLPKELFORKMLAANKF 539
QY 570 NTGLTLRQIVLSKVDQSLHTNTSLDAASEYAKYCEI---LGVAAATPGTNMPA-TFGHL 625
DB 540 QRGMFLVRQMEFALFDMYIYSESDECRKNWQVLDVSRKEAVIQPPEYNFANSFGHI 599
QY 626 -AGGYDQYGYLWSEVFSMDMFYSCFKKKGIMNPEVGMKYRNLLKPGSLDGMMLHN 684
DB 600 FAGGYAGYISYAWAEVLSTDA-YAAFEESDDV-AATGKRFWQEIILAVGGRSRAESFKA 657
QY 685 FLKREPNOKAFLMSRG 700
DB 658 FRGPREPSIDALLRQSG 673

RESULT 4
US-09-806-866-57
; Sequence 37, Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Massignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manroza
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scariato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
```

```
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; US-09-806-866-57

Query Match          19.1%; Score 702; DB 5; Length 678;
Best Local Similarity 28.4%; Pred. No. 2.2e-40;
Matches 194; Conservative 126; Mismatches 302; Indels 60; Gaps 19;

QY 55 EQIKTRTEELIVQT--KQYDVAVGMGLGIEEVT--YENCLOALADVEVKYIVERT-----ML 106
DB 16 DQIKTEDIKPAQTAIAEAREQIAAQAQTHGTGWANTVEPLTG-----ITERVGRWGVV 70
QY 107 DFPQHVSSDKVEVRAASTADKRLSRFDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYL 166
DB 71 SHLNSVADPELRVAVNELMPEITVFTFTEIGQDIELYNRFKTIKNSPEFDTLSPAQKTLL 130
QY 167 EKSIRMGKRNGLHLPEOVQNEIKSMKRSELCIDFNKN-LNEDDTFLVF--SKAELGAL 223
DB 131 NHDLRDVLVSGAELPERQAEALQTEGAQLSAKFSQNVLDATDAFGYFDDAAPLAGI 190
QY 224 PDD-----FIDSLEKTDQDKYKITLKYPHYFPVWKKCCIPETRRMEMAFNTRCKE----- 274
DB 191 PEDALAMFAAAQSEKTKYKIGLQIPHYLAVIQVAGNRELREQIYRAVTVTRASELSDDG 250
QY 275 --ENTILQOLLPLRTKVAKLGLGYSTHADFLVEMNTAKSTSRVTAFLDLSOKLPLGEA 332
DB 251 KFDNTANIDRTLLENALQTAALLGFKNYAELSLATKMDTPEQVLFNHLDRARRAKPYAEK 310
QY 333 E-----REFILNLKKKCKDRGFYDQKINAMDLYVYMTQTEELKYSIDQEFLEKYP 384
DB 311 DLAEVKAFARE-SLNL-----ADLPQWDLGVASEKLEAKYAFSETEVKYKF 356
QY 385 PIEVTVTEGLNLYQELLGLSFQMTDAHWNKSVTLTYVKDKATGEVLGQFVLDLYPREG 444
DB 357 PVGKVLNGLFAQIKKLYGIGFTEKT--VPVWHKDVRYFELQ--NGETIGGVYMDLYAREG 413
QY 445 KYNHAACFGLQPCLLPGSRMMAVAALVNFPSQVAGRPSLLRHDEVRTTFHEFGHVMH 504
DB 414 KRGGAMNDYKGRRRFSDGTQLPTAYLVNCFNFPVPGVKGEARLSHDEITLTFHETGHGLH 473
QY 505 QICAQTDDFARFSGTN-VETDFVEVPSQMLENNWVDVSLRLSKHYKDGSPITADDLLEKL 563
DB 474 HLLTOVDDELGVSGINGVENDAVELPSQMFNFVWYNVLAQMSAHEETGVPLPKELFDM 533
QY 564 VASRLVNTGLTLRQIVLSKVDQSLHTNTSLDAASEYAKYCEI---LGVAAATPGTNMPA 620
DB 534 LAAKNFQRCNMFVRQMEFALFDMYIYSEDDEGRKNWQVLDVSRKKAIVQPPENRFA 593
QY 621 -TFGHL-AGGYDQYGYLWSEVFSMDMFYSCFKKKGIMNPEVGMKYRNLLKPGSLDGM 678
DB 594 LSFGLHIFAGYSAGYISYAWAEVLSADA-YAAFEESDDV-AATGKRFWQEIILAVGGRSRA 651
QY 679 MDMLHNLKREPNOKAFLMSRG 700
DB 652 AESFKAFRGREPSIDALLRHS 673
```

```
RESULT 5
US-09-806-866-55
; Sequence 55; Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manrosa
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-806-866-55

Query Match 18.98; Score 694; DB 5; Length 678;
Best Local Similarity 28.28; Pred. No. 7.7e-40;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQIKTRTEELIVQT--KQYDAVGMGLGEEVT-YENCLOALADVEKVIYVERT-----ML 106
DB 16 DQIKTEDIKPALQTAIAEAREQIAAIAKQATHGTWANTVEPLTG-----ITRVGRINGVV 70
QY 107 DFPQHVSSDKKEVRAASTPADKRLSRFDIEMSMRGDIFERIVHLOETCDLGKIKPEARYL 166
DB 71 SHLSNVDTPELRAAYNELMPEITVFFTEIGQDIELYNREKTIKNSPEFDTLSHAQTKL 130
QY 167 EKSKIMGRNGLHLPQEQVONIEKSMKKRMSELCDIFNKN-LNEDDTFLVF--SKAELGAL 223
DB 131 NHDLRDFVLSGAELPPEQQAELALQTEGAOLSAKFSQNLVDATDAFGIYFDDAAPLAGI 190
QY 224 PDD-----FIDSLKTDDBKXITIKYHYFVPMKKCCIPETRRRMEMAFNTRCKE----- 274
DB 191 PEDALAMFAAAQSECKGYKIGIQIPHYLAVIQYADNRKIREQIYRAYVTRASELSDG 250
QY 275 --ENTIILOQLPLRTKVAKLLGYSTHADVFLEMMNTAKSTSRVTAFLDLSQKLKPLGEA 332
DB 251 KFDNTANDITRENALQTAQLGLGFKNYAELSLSATKMDATPEQVLNLFHDLARRAKPYAEK 310
QY 333 EREFTLNKKKECKDRGEYDCKNANDLYYYMTQTTELKYSIDQEFLEKFPFIEVTVTEG 392
DB 311 D--LAEYKAFARESLG---ADLPQNDLVAGELKREAKYAFSETEVKYFPVPGKVLNG 364
QY 393 LLNTYQELLG--SFEQMTDAHVWNKSVTLTYVKDKATGVLGQFVLDLYPREGKYNHAACF 452
DB 365 LFAQIKKLYIGTFEKT--VPVHHKDVRYFEIQQ--NGETIGGVTDMLYARGGKRGGAWN 421
QY 453 GLQPGCLLPDGRMMAVAALVNFNSQPVAGRPSLLRHDEVRTYFHEFGVHMHOICAQTFD 512

US-09-806-866-61
; Sequence 61; Application US/09806866
; GENERAL INFORMATION:
; APPLICANT: Frazer, Claire M.
; APPLICANT: Hickey, Erin
; APPLICANT: Peterson, Jeremy
; APPLICANT: Tettelin, Herve
; APPLICANT: Venter, J. Craig
; APPLICANT: Masignani, Vega
; APPLICANT: Galeotti, Cesira
; APPLICANT: Mora, Manrosa
; APPLICANT: Ratti, Giulio
; APPLICANT: Scarselli, Maria
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagratia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisseria Genomic Research
; FILE REFERENCE: CHIR-0313
; CURRENT APPLICATION NUMBER: US/09/806,866
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: PCT/US00/05928
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: PCT/US99/25373
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 60/103,794
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: 60/132,068
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-806-866-61

Query Match 18.98; Score 694; DB 5; Length 678;
Best Local Similarity 28.28; Pred. No. 7.7e-40;
Matches 190; Conservative 131; Mismatches 309; Indels 44; Gaps 18;

QY 55 EQIKTRTEELIVQT--KQYDAVGMGLGEEVT-YENCLOALADVEKVIYVERT-----ML 106
DB 16 DQIKTEDIKPALQTAIAEAREQIAAIAKQATHGTWANTVEPLTG-----ITRVGRINGVV 70
QY 107 DFPQHVSSDKKEVRAASTPADKRLSRFDIEMSMRGDIFERIVHLOETCDLGKIKPEARYL 166
DB 71 SHLSNVDTPELRAAYNELMPEITVFFTEIGQDIELYNREKTIKNSPEFDTLSHAQTKL 130
QY 167 EKSKIMGRNGLHLPQEQVONIEKSMKKRMSELCDIFNKN-LNEDDTFLVF--SKAELGAL 223
```

```
Db 131 NHDRLRDLVLSGABLPPEQQAELAKLQTEGAQLSAKFSQNVLDATDAFGIYFDAAAPLAGI 190
QY 224 PDD- ---FIDSLEKTDKDKYKTLKYPHYFPYMKKCCIPETRRMEMAENTRCKE----- 274
Db 191 PEDALAMFAAAQSEKGTGKGLQIPHYLAVIQYADNRKLRQEIYRAXVTRASELSDDG 250
QY 275 --ENTIILOQLPLRTKVAKLKCYSTHADPVLNMTAKSTSRVTAFLDLDLSOKLKLPLGEA 332
Db 251 KPDNTANIDRTLENALQAKLLGFKNYAELSLATKNADTPQVNLNHLARRAKPYAEK 310
QY 333 ERFILNLKKCKDRGFYDVKINAWDLYYYMTQTEELKYSIDQEFLEKYPPIEVVTEG 392
Db 311 D---LAEVKAFARESLGL---ADLPQWDLGYAGEKLRKAYAFSETEVKKYPVGVKVLNG 364
QY 393 LLNTYQELLGLSPEQMTDAHVWNKSVTLTVKDKATGEVLGQFYLDLYPREGKYNHAACF 452
Db 365 LFAQIKKLYGIGTEKT-VPVWHKDVRYFELQO--NGETIGGVYMDLYAREGKRGGAWN 421
QY 453 GLQPGCLLPDGRMMAVAALVNFSPVAGRPSLLRHDEVRTYFHEFGHVMHOICAQTOF 512
Db 422 DYKGRFRSDGTQLQTLAYLVCNFTPPVGKKEARLSHDELLTLFHTGHLHLLTQVDE 481
QY 513 ARPSGTN-VETDFEVPQSOMLENWVDLSRLSKHYKDGSPIDADDLLEKLVASRLVNT 571
Db 482 LGVSGINGVEMDAVELPSQFMENFVWYNVLAQMSAHEETGVPLPKELFDKMLAAKNFOR 541
QY 572 GLTLRQIVLSKVDQSLHNTSLDAASEYAKYVCSEI---LGVNAATPGTNMPA-TFGHL-A 626
Db 542 GMFLVRQMEFALFDMYIYSEDDEGRLLKNQOVLDSVRKEVAVVVRPPEYNRFANSFGHIFA 601
QY 627 GYDGOYGYGLWSEVFSMDYFCFKKEGIMNPEVGMKYRNLLIKPGSLDGMDLHNL 686
Db 602 GGSAGYYSYANAEVLSADA-YAAFEESDDV-AATKRFWQEIILAVGGRSRAESKAFR 659
QY 687 KREPNOKAFILMSRG 700
Db 660 GREPSIDALLRHSG 673

RESULT 7
US-10-015-127-11537
; Sequence 11537, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
; FILE REFERENCE: 38-10(15806)B
; CURRENT APPLICATION NUMBER: US/10/015,127
; PRIOR FILING DATE: 2001-10-29
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 14357
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Sphingomonas elodea
US-10-015-127-11537

Query Match 17.5%; Score 641.5; DB 6; Length 678;
Best Local Similarity 28.1%; Pred. No. 3,1e-36;
Matches 189; Conservative 122; Mismatches 327; Indels 35; Gaps 16;

QY 52 LSPEQIKRTEELIVQTKQYDVGMLGIEEYVYENCALQADLVVEKYIVERTMLDFPOH 111
Db 17 IHPDQIAPALDAVIAEHAQAAVERV--IASESRFDDVWMPLEAETAIDALWSTVSHLRG 74
QY 112 VSSDKVEVRAASTADKRLSRFDIEMSRGDIPIERIVHLQETCDLGIKQKEA-RRYLEKSI 170
Db 75 VADTPELRAAHAGAGARLTIENQNLAVNQNALVEVLVATATPDFA-ARPOADRAAYEHMV 133
QY 171 KMCKRGNLHLPQEQVQNEIKSMKRMSELCTDFNKNLNEDDTFLVFSKAE-----LGALPD 225
```

```
Db 134 RDRFLSGVALDAEAKARFAAASVSELSTLSTGEG-NAVLDATDAWFEHIEDPALLAGISPA 192
QY 226 D---FIDSLEKTDKDKYKTLKYPHYFPYMKKCCIPETRRMEMAENTRCKE----- 274
Db 193 PEDALAMFAAAQSEKGTGKGLQIPHYLAVIQYADNRKLRQEIYRAXVTRASELSDDG 250
QY 275 --ENTIILOQLPLRTKVAKLKCYSTHADPVLNMTAKSTSRVTAFLDLDLSOKLKLPLGEA 334
Db 253 DNGPRIAALLELRREAQQLGPPDPVANSLATKMPAAGEVLAFRLDGLRRARPAQAQAE- 311
QY 335 EFLINLKKCKDRGFYDVKINAWDLYYYMTQTEELKYSIDQEFLEKYPPIEVVTEGL 394
Db 312 --FAELSAFAARELGD---TLQPDVDPFASDRRLAARYAVDQEVRAHFPPVERVIAGWQ 366
QY 395 NYIOELLGLSFEQMTDAHVWNKSVTLTVKDKATGEVLGQFYLDLYPREGKYNHAAACFGL 454
Db 367 ALLARLFGVRLVARPDVAVYHPDACYIDYVD-AEGVVIAGVYTDLHARAGKRSAGWMAQA 425
QY 455 QPGCLLPDGS-RMMAVAALVNFSPVAGRPSLLRHDEVRTYFHEFGHVMHOICAQTOF 513
Db 426 RP--RUNDGNVRPVPAYLYVCNFPKPTGEGSLLSHPEVVTLLHETGCHLHLFTVRDRP 483
QY 514 RFSGTN-VETDFEVPQSOMLENWVDLSRLSKHYKDGSPIDADDLLEKLVASRLVNTG 572
Db 484 NIAGTSGFEMDAVELPSQLMEDFAWDROVLRGMSGHATGAPLPDVLFDKLIARREFLAG 543
QY 573 LTLRQIVLSKVDQSLHNTSLDAASEYAKYVCSEIILGVNAATPGTN-MPATFGHL-AGGYD 630
Db 544 MALVRQIEFALFDLLHLGTMGSDPMKVIKAVRDEAVAVVRPPEWHRFPHAFSHIFAGGYA 603
QY 631 GOYGYGLWSEVFSMDYFCFKKEGIMNPEVGMKYRNLLIKPGSLDGMDLHNLKREP 690
Db 604 SGYISYLAWEVLAADGFGQ-FAEAGLIDRATADRFRDEVLRGASRPAAESFRAFGRDP 662
QY 691 NOKAFLMSRGLHA 703
Db 663 DPOAMLLRHGLTA 675

RESULT 8
US-09-897-516-8169
; Sequence 8169, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 8169
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-8169

Query Match 15.0%; Score 550; DB 5; Length 691;
Best Local Similarity 26.2%; Pred. No. 6,1e-30;
Matches 163; Conservative 117; Mismatches 267; Indels 74; Gaps 16;

QY 122 STBADKRLSRFDIEMSRGDIPIERIVHLQETCDLGIKQKEAPARRYLEKSKMKGRNGLHLP 181
Db 100 SEWSPKLAAMNDEIMLSKLFNRKTIYQOETNLDSRLVEIVYKQFELAGANLS 159
```



Db 227 FSDGTLQPTAYLVCNFPAPVGGREARLSDHEILILFHETGHLHLHLLTQVDELGVSGIN 286  
QY 520 -VETDFEVEPSQMLENNWVDVDSIRLSKHVKDGSPIADDDLEKLVASRLVNTGLLTLRQ 578  
Db 287 GVXWDVAVELSPQMFNFVWYNVLAQSAHEETGVPLPKELXDKXLAANKFQXMFXYRQ 346  
QY 579 IVLSKQDSQSLTHTSLDAASEYAKYCSSEI---LGVAAATPGTNNMPA-TFGHL-AGGYDGGY 633  
Db 347 XEPALFDMYIYSEDDGRLKNWQVLDVSRKKVAVIOPPEYNRFALSFCHIFAGGYSAAX 406  
QY 634 YGLWSEVFMDFSCFKKEGTMNFEVGMKYNRLILKPGSGSLDGMMDLHNFLLKREPNOK 693  
Db 407 YSYAWAEVLSADA-YAAFEESDDV-AATGKRWFQETILAVGXSRGSAESKAFRGREPSID 464  
QY 694 AFLMSRG 700  
Db 465 ALLRHSG 471  
RESULT 10  
US-09-806-866-59  
; Sequence 59, Application US/09806866  
; GENERAL INFORMATION:  
; APPLICANT: Frazer, Claire M.  
; APPLICANT: Hickey, Erin  
; APPLICANT: Peterson, Jeremy  
; APPLICANT: Tettelin, Hervé  
; APPLICANT: Venter, J. Craig  
; APPLICANT: Masignani, Vega  
; APPLICANT: Galeotti, Cesira  
; APPLICANT: Mora, Manroza  
; APPLICANT: Ratti, Giulio  
; APPLICANT: Scarselli, Maria  
; APPLICANT: Scarlato, Vincenzo  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagratia  
; APPLICANT: Grandi, Guido  
; TITLE OF INVENTION: Neisseria Genomic Research  
; FILE REFERENCE: CHIR-0313  
; CURRENT APPLICATION NUMBER: US/09/806,866  
; CURRENT FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: PCT/US00/05928  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: PCT/US99/25373  
; PRIOR FILING DATE: 1999-10-08  
; PRIOR APPLICATION NUMBER: 60/103,794  
; PRIOR FILING DATE: 1998-10-09  
; PRIOR APPLICATION NUMBER: 60/132,068  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 59  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-09-806-866-59

Query Match 12.9%; Score 475; DB 5; Length 491;  
Best Local Similarity 28.0%; Pred. No. 5.4e-25;  
Matches 137; Conservative 88; Mismatches 222; Indels 42; Gaps 12;  
QY 56 QIKTRTEELIVQTKYVDVAVGMLGIEVET---YENCLQALADVEVKYIVERT-----ML 106  
Db 17 QIKTEDIKPAVQT-AIAEARGQAAVKAQTHGTWANTVERLTG-----ITERVGRWGVV 70  
QY 107 DFPQHVSSDKVEAASAEADKRLSDIEMSRGDIFFERIVHLQETCDLGKIKPEARRYL 166  
Db 71 SHLNSVVDTEPRLAVYNELMPETVFFTEIGQDIELYNNRFTIKNSPEATLSPAOKTKL 130  
QY 167 EKSINKGRNGLHLPEQVONEIKSMKKRMSELICIDFNKN-LNEDDFTFLVFP--SKAELGAL 223  
Db 131 DHDLRDFVLSGAELEPPERQAEALAKLQTEGAQLSAKFSQNVLDATDAFGIYFDDAAPLAGI 190

QY 224 PDD---FIDSEKTDKDKYKITLKYPHYPPVMKKCCIPETRRMEMAFNTRCKE----- 274  
Db 191 PEDALAWFAAAQSEKGTGKIGLQIPIHLAVIQYAGNRELREQIYRAYVTRASELSNDG 250  
QY 275 --ENTTILOQLPLRTKVAKLGLSYTHADFLVEMNTAKTSRYTAFIDDLQSOKLPIGEA 332  
Db 251 KFDNTANIDRTLENALKTAKLGLFKNYAELSLATKNADTPEQVLNLFHDLARRAKPYAEK 310  
QY 333 EREFIILNKKKECKDRGFEDGKIN--AMDLYYMTQTEELKYSIDOEFLKEYFPFIEVVT 390  
Db 311 D-----LAEYKAFAREHLGLADPQWDLVSAGEKLEAKYAFSETEVKYPFGVKVL 362  
QY 391 EGLNTYQELGLGSEFQMTDHAHVWNSVTLYTVKDKATGEVLGQVLYLDLYPREGKYNHAA 450  
Db 363 AGLFAQIKKIDYIGGFAEKT--VPVWHKDRYFELQQ--NGKTIQGVYMDLYFARGKRGGA 419  
QY 451 CFGLOPGCLLPDGRMMAVAALVNVFNSQPVAGRPSSLRHDEVRTYFHEFGHVMHOICAO 510  
Db 420 MNDYKGRRRFADGTQLQPTAYLVCNFPAPVGGKEARLSHDEILTLFHETGHLHLHLLTQ 479  
QY 511 DPARFSGTN 519  
Db 480 DELGVSGIN 488  
RESULT 11  
US-09-826-734-44  
; Sequence 44, Application US/09826734  
; GENERAL INFORMATION:  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Mishra, Vishnu S.  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Zerhusen, Bryan D.  
; APPLICANT: Kekuda, Ramesha  
; TITLE OF INVENTION: Novel Polynucleotides and Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-754  
; CURRENT APPLICATION NUMBER: US/09/826,734  
; CURRENT FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 60/195,576  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 270  
; SOFTWARE: FastSeq for Windows version 4.0  
; SEQ ID NO 44  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-826-734-44

Query Match 10.2%; Score 373; DB 5; Length 71;  
Best Local Similarity 100.0%; Pred. No. 3.7e-19;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 182 EQVONEIKSMKKRMSELICIDFNKNLNEDDTFLVFSKAEALGALPDDFIDSLEKTDKDKYKI 241  
Db 1 EQVONEIKSMKKRMSELICIDFNKNLNEDDTFLVFSKAEALGALPDDFIDSLEKTDKDKYKI 60  
QY 242 TLKYPHYFPVM 252  
Db 61 TLKYPHYFPVM 71

RESULT 12  
US-10-015-127-11338  
; Sequence 11338, Application US/10015127  
; GENERAL INFORMATION:  
; APPLICANT: Bower, Stanley G.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof

FILE REFERENCE: 38-10(15806)B  
CURRENT APPLICATION NUMBER: US/10/015.127  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: US 60/252,455  
PRIOR FILING DATE: 2000-11-22  
NUMBER OF SEQ ID NOS: 14357  
SEQ ID NO 11338  
LENGTH: 251  
TYPE: PRT  
ORGANISM: Sphingomonas elodea  
US-10-015-127-11338

Query Match 9.2%; Score 338.5; DB 6; Length 251;  
Best Local Similarity 31.9%; Pred. No. 5e-16;  
Matches 75; Conservative 45; Mismatches 108; Indels 7; Gaps 3;  
Qy 475 NFSQVAGRPRLRLHDEVRTTFEFGHVMHQICATDFARFSGTNVETDFEVPQSMLEN 534  
Db 7 NPTKPAQGPALISDDVTTFEFGHALHGLFANQTPSVSGTNVARDFAEFPQFNEH 66  
Qy 535 VWMDVDSLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRLQIVLSKVDQSLHTNTSL 594  
Db 67 WALDPKVLPHYAVNTKDGSVIPQALVDKIKRAGTNSGYSGEALAAEMDMWSHLAA 126  
Qy 595 DAASEYAKYCSBILGVAATPGTNMP-----ATFGHLAG-GYDGOYGYLWSEVFSMDMFY 648  
Db 127 DGKQDADAEQKALAAATGLDVTDPYPRYSSVFLHWGYSAGYAYMTKMLSANAF- 185  
Qy 649 SCFKKEGTMNPEVMGMYNRLILKPGSGLDGMDMLNFKRPNQKAFMLSGRLHA 703  
Db 186 NWFQHGGMTRANGORFREMVLKSGHTEDYAPMFRFNGADFPQVAPLKLKDLGLNA 240

RESULT 13  
US-09-708-427-15046  
; Sequence 15046, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15046  
; LENGTH: 1144  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1144  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1144  
; OTHER INFORMATION: Ceres Seq. ID 1828629  
US-09-708-427-15046

Query Match 3.5%; Score 130; DB 5; Length 1144;  
Best Local Similarity 18.2%; Pred. No. 0.86;  
Matches 111; Conservative 84; Mismatches 194; Indels 220; Gaps 24;  
Qy 60 RTEELIVOTKQVDAVGML-----GIEEVTYENCLOALADVEVKYI 100  
Db 520 KIEELSVANESLADNVDTLQISVQESKDLKEREVAYLKIEELSVAN--ESLVDKETKL- 576  
Qy 101 VERTMLDPFQVSSD-KEVRAASTEADKRLSRFDIEMSRGDIFFERIVHLQETCDLGR-I 158  
Db 577 -----QHIDQAEELRGREASHLKKIEELSK-----NENLVNVMNQNIABESKDL 624  
Qy 159 KPEARRYLEKSIKMGKRNGLHLPQEQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKA 218

Db 625 REREVAYLKIKIDELSTANGT-LADNVTN-----LQNISE-----ENKELRERETILLKKA 674  
Qy 219 ELGALPDDFIDSLEKTDODDKYKITLKYPHYFPVMKKCCIPETRRRMEAFNTRCKEENT 278  
Db 675 ELSNELSLVDKASKL-----QTVVQE-----NEELRERETA 706  
Qy 279 ILQQLPLRTKVAKLLGYSTHADFVLEMTAKTSRVTAFDDLSQKLPGLGEAREFILL 338  
Db 707 YLKKI-----EELSKLHEILSDQETKLQ-----IS 731  
Qy 339 NLKKECKDRGFEYDKINANDLYYYTQTTEELKYSIDQEFLEKVFPIEVTEGLNTYQ 398  
Db 732 NHEKEELKER-----ETAYLKKIEELSKYQEDLLNKEN 764  
Qy 399 ELLGLSEF---OMTDAHVMNKSVTLYTVKDKATGEVLGQFVLDLYPREGYNHAAFCG- 453  
Db 765 ELHGMVVEIDLRSDSLAOKKI-----EELSNNASLLIKENELQAVVCENE 812  
Qy 454 -----LQPGCLLPDGSRRMAYALVNFVQSVAGRP 484  
Db 813 ELKSKQVSTLKTIDELSDLKQSLIHKEKELQ--AAIVENEKLAEEAALSQRIEELTNLK 870  
Qy 485 SLL--RHDEVITYEHE-----FGHVMHQICATDFARFSGTNVETDFVEVP 528  
Db 871 QTLIDKQNELQGVFHENEELKAKAEASSLKIDELHL-----LEQSWLEKESFQRT 922  
Qy 529 SOMLENVMDVDSLRRLSKHYKDGSPADDDLEKLVASRLVNTGLTLRLQIVLSKVDQ-S 587  
Db 923 QENLELKTQDALAAKKI-----EELSKLIESLLEK-----ETELKCREAAALEKMEEPS 971  
Qy 588 LHTNTSLDA 596  
Db 972 KHGNSLNS 980

RESULT 14  
US-09-708-427-15045  
; Sequence 15045, Application US/09708427  
; GENERAL INFORMATION:  
; APPLICANT: N. ALEXANDROV et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID  
; FILE REFERENCE: 2750-1243P  
; CURRENT APPLICATION NUMBER: US/09/708.427  
; CURRENT FILING DATE: 2000-11-09  
; NUMBER OF SEQ ID NOS: 85364  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15045  
; LENGTH: 1304  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: 1..1304  
; OTHER INFORMATION: Xaa is any amino acid  
; NAME/KEY: misc.feature  
; LOCATION: 1..1304  
; OTHER INFORMATION: Ceres Seq. ID 1828628  
US-09-708-427-15045

Query Match 3.5%; Score 130; DB 5; Length 1304;  
Best Local Similarity 18.2%; Pred. No. 1;  
Matches 111; Conservative 84; Mismatches 194; Indels 220; Gaps 24;  
Qy 60 RTEELIVOTKQVDAVGML-----GIEEVTYENCLOALADVEVKYI 100  
Db 680 KIEELSVANESLADNVDTLQISVQESKDLKEREVAYLKIEELSVAN--ESLVDKETKL- 736  
Qy 101 VERTMLDPFQVSSD-KEVRAASTEADKRLSRFDIEMSRGDIFFERIVHLQETCDLGR-I 158  
Db 737 -----QHIDQAEELRGREASHLKKIEELSK-----NENLVNVMNQNIABESKDL 784

Qy	159	KPEARRYLESIKMGKRNGLHLPEQVQNEIKSMKRMSELCDIFNKNLNEDDTFLVFSKA	218
Db	785	REEREVAYLKKIDELSTANGT--LADNVN-----LQNISE---ENKELRERETTLKKAE	834
Qy	219	ELGALPDFIDSLSEKTDQDKYKILKYPHYFVVMKKCCIPETRRRMEWAFNTRCKEWTI	278
Db	835	ELSENESLVDKASKL-----QTVQE-----NEELRERETA	866
Qy	279	ILQQLPLRTKVAKILGLGYSTHADFVLEMNTAKTSRVTAFLDDLSOKLKLPLGEAREFTL	338
Db	867	YLKKI-----EELSKLHEILSDQETKLQ-----IS	891
Qy	339	NLKKKECKDRGFYDGKINAWDLYYMTQTEELKYSIDQEFKELYFIEVWTEGLLNTYQ	398
Db	892	NHEKEELK-----ETAYLKKIEELSKVQEDLLNKEN	924
Qy	399	ELLGISE---QMTDAHVNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHAACFG-	453
Db	925	ELHGMVVEIEDLRSKDSLAQKKI-----EELSNFNASLLIKENELQAVVCNE	972
Qy	454	-----LQPGCLLPDGSRMVAALVWNFSQPVAGRP	484
Db	973	ELSKQVSTLKTIDELSDLKOSLIHKEKELQ--RAIVENELKKAERALSQRIBELTNLK	1030
Qy	485	SLL--RHDEVITYPHE-----EGHVMHQICAQTDFAFSGTNNVETDFVEYP	528
Db	1031	QTLTDKQNELQGVFENEBELKAKEASSLKKIDELH-----LEQSWLEKESFQVYT	1082
Qy	529	SOMLENWVDVSLRRSLKHKYKDSPIADDDLEKLIVASRLVNTGLTLRQIVLSKVDQ-S	587
Db	1083	QENLELTKQDALAAKKI---EELSKLESLEK-----ETELKCREAAALEKMEERPS	1131
Qy	588	LHTNTSLDA	596
Db	1132	KHGNSELNS	1140

RESULT 15  
 US-09-708-427-15044  
 ; Sequence 15044, Application US/09708427  
 ; GENERAL INFORMATION:  
 ; APPLICANT: N. ALEXANDROV et al.  
 ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
 ; TITLE OF INVENTION: THEREBY  
 ; FILE REFERENCE: 2750-1243P  
 ; CURRENT APPLICATION NUMBER: US/09/708,427  
 ; CURRENT FILING DATE: 2000-11-09  
 ; NUMBER OF SEQ ID NOS: 85364  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 15044  
 ; LENGTH: 1313  
 ; TYPE: PRT  
 ; ORGANISM: Arabidopsis thaliana  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1..1313  
 ; OTHER INFORMATION: Xaa is any amino acid  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1..1313  
 ; OTHER INFORMATION: Ceres Seq. ID 1828627  
 ; US-09-708-427-15044

[illegible]

Search completed: January 14, 2002, 19:00:16  
Job time: 340 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: January 14, 2002, 18:57:51 ; Search time 1811.41 Seconds

(without alignments)  
12546.757 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcgccggcgccctttt.....gagggcctgcctgcgtcgtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estom: \*  
5: em\_estpi: \*  
6: em\_estba: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_hic: \*  
10: gb\_estl: \*  
11: gb\_est2: \*  
12: gb\_hic: \*  
13: gb\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_inv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rod: \*  
20: em\_gss\_vrt: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	622	29.4	625	10	AI952756
2	611	28.9	789	11	BI335570
3	570	27.0	579	11	BI180236
4	506.8	24.0	902	11	BI144134
5	479.4	22.7	745	11	BI144134
6	475.6	22.5	619	10	AW742508
7	472.6	22.3	745	11	BI144134
8	455.6	21.5	559	10	BE311015
9	431.4	20.4	509	10	AW786845
10	419.8	19.8	570	10	AW732454
11	416.8	19.7	511	10	AI386253
C 12	390.6	18.5	757	10	AW003054

C 13	376.8	17.8	680	11	BG380774
14	373.2	17.6	675	11	BF719641
15	358.2	16.9	417	10	AA152501
C 16	358	16.9	680	10	BE630187
17	344.8	16.3	447	10	AA066229
C 18	341.4	16.1	750	10	AI326417
19	331.4	15.7	366	10	AW859092
20	331.4	15.7	889	11	BG245790
21	328.8	15.5	560	10	BE627164
22	324.6	15.3	654	11	BF535533
23	318.8	15.1	984	11	BF974442
24	315.2	14.9	660	11	BI393320
25	306.4	14.5	759	11	BG967757
26	302.4	14.3	517	10	AI195498
27	299.4	14.2	818	10	AL521398
28	292.8	13.8	857	11	BG758404
29	289	13.7	765	10	AI322656
30	288.8	13.7	956	10	AL553474
31	280	13.2	776	11	BG178482
32	278.8	13.2	777	11	BI327686
C 33	272.2	12.9	462	10	AI195785
34	265.8	12.6	902	11	BF664439
35	264	12.5	717	11	BF528020
36	261.6	12.4	907	11	BE875429
37	258.6	12.2	901	10	AL551921
38	258.2	12.2	729	11	BG260342
39	257.2	12.2	725	11	BI194940
C 40	256.8	12.1	382	10	AW742681
C 41	255	12.1	706	11	BF222737
42	255	12.1	759	11	BG686225
43	252.2	11.9	530	11	BI314174
44	251.2	11.9	949	11	BG180024
45	250.6	11.8	748	11	BG972692

#### ALIGNMENTS

RESULT 1  
AI952756/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AI952756 625 bp mRNA EST  
wr50b06.x1 NCI-CGAP\_Utl Homo sapiens cDNA clone IMAGE:2491091 3'  
similar to SW:NEUL\_PIG Q02038 NEUROLYSIN PRECURSOR ;, mRNA  
sequence.

AI952756  
AI952756.1 GI:5745066

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 625)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1807 Std Error: 0.00

Seq primer: 40UP from Gibco

High quality sequence stop: 424.

Location/Qualifiers

1. 625

/organism="Homo sapiens"

/db\_xref="taxon:9606"

FEATURES

source

```
/clone="IMAGE:2491091"
/clone_lib="NCI_CGAP_Utl1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT      175 a  155 c  134 g  158 t      3 others
ORIGIN

Query Match.          29.4%; Score 622; DB 10; Length 625;
Best Local Similarity 99.5%; Pred. No. 3e-142;
Matches 622; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1333 aaatacaatcatgaggcctgttgggtccagctggtcgtccttctgctgtaggaagc 1392
Db 625 AAATACAATCATGCGGCTGTTCGGTCTCCAGCTGGCTGCCCTCTCNCNTGATGGAAGC 566

Qy 1393 cggatgatggcagtggtgctcctcgtggtgaactctcacagccagtggcaggctgctcc 1452
Db 565 CGGATGATGGCAGTGCGTCCCTCGTGGTGAACCTCTCACAGCCAGTGCGCAGGTCGTC 506

Qy 1453 tctctcctgagacagcagtgagtgacttacttcatgagttgttcacgtgatgcatt 1512
Db 505 TCTCTCCGACACAGCAGGAGTGAGGACTTACTTTTCAATGATTTGGTCAGCTGATGCT 446

Qy 1513 cagatttgcacagactgatttgcagatttagcggaacaaatgtggaactgacttt 1572
Db 445 CAGATTTGTGCACAGACTGATTTTGCAGGATTTAGCGGAACAATGTGGAACCTGACTTT 386

Qy 1573 gtagagtgccatgcgaaatcgtgaaattgggtgtgggacgtcgattccctccgaga 1632
Db 385 GTAGAGTGCCATCGCAAAATGCTTCAAAATTTGGGTGTGGGACGTCGATTTCCCTCCGAGA 326

Qy 1633 ttgtcaaaacattataagatggaagccctatttcgacagctctgcttgaacttggtt 1692
Db 325 TTGTCAAAACATTATAAGATGGAAGCCCTATTTCGACAGCATCTGCTTGAACACTTGT 266

Qy 1693 gcttctaggctggttcaaacacaggttcttctgaccctgcgcagattgttttgagcaagtt 1752
Db 265 GCTTCTAGGCTGGTCAACACAGGCTCTTCTGACCTGCGCCAGATTGTTTTCGACANAGTT 206

Qy 1753 gatcagctctctcaccacacacacacacacacacacacacacacacacacacacacac 1812
Db 205 GATCAGTCTCTTCATACCAACACATCGCTGGATGCTGCAAGTGAATATGCCAAATACTGC 146

Qy 1813 tcagaaatattaggagttgcagctactccaggcacaaatagccagctacacctttggacat 1872
Db 145 TCAGAAATATTAGAGTTGCAGCTACTCCAGGCACAAATATGCCAGCTACCTTTGGACAT 86

Qy 1873 ttggcaggggatacagatggccaattatttgatatcttttggagtgaaagtattttccatg 1932
Db 85 TTGCGAGGGGATACGATGGCCAATATTATGGATATCTTTGGAGTGAAGTATTTTCCATG 26

Qy 1933 gatattgtttacagctgttttaaaa 1957
Db 25 GATATGTTTTACAGCTGTTTTAAAAA 1

RESULT 2
BI335570 LOCUS BI335570 789 bp mRNA EST 30-JUL-2001
DEFINITION 602998590F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5140583 5',
mRNA sequence.
ACCESSION BI335570
VERSION BI335570.1 GI:15020227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 789)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11346 row: d column: 24
High quality sequence start: 33
High quality sequence stop: 789.
Location/Qualifiers
1..789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5140583"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/Note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

BASE COUNT      226 a  177 c  211 g  175 t
ORIGIN

Query Match          28.9%; Score 611; DB 11; Length 789;
Best Local Similarity 99.0%; Pred. No. 1.5e-139;
Matches 625; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 atgatcgccggtgcttcttggctgctggaagcctccgagagtggtggtccaggatt 60
Db 160 ATGATCGCCCGTGCCTTTTGGCTGTGGGAAGCCTCCGACAGTTGGTGTCCAGGATT 219

Qy 61 ttactcagaatgacgttaggaagaagtgtctctctctctcaggaatgtcttccat 120
Db 220 TTACTCAGATGACGTAGGAAGAGAGTGTCTCTCTTCAGGCAATGCTTCTCTAT 279

Qy 121 acttgctggcagaataatgttttaagatgggatcttccacagagcaaatataaacaaga 180
Db 280 ACTGTGCTGGCAGAAATGTTTTAAGATGGGATCTTTCCACAGACAAATTAACAAGA 339

Qy 181 actgagagctcattgtgcagaccacacacaggtgtacgatgctgtggaatgctcggtatt 240
Db 340 ACTGAGGAGCTCATTGTGCAGACCAACAGGTGTACGATGCTGTGGATGCTCGGTATT 399

Qy 241 gaggaagttaactacgagaactgtctgcagcactgcgcagatgtagaagtaaatatata 300
Db 400 GAGGAAGTAACTTACGAGAACTGTCTGCAGCACTGGCAGATGTAGAAGTAAAGTATATA 459

Qy 301 gtggaaggagcattgctagactttcccccagcatgtatctctgcacaaagaagtcagagca 360
Db 460 GTGGAAGGAGCATTGCTAGACTTTTCCCCAGCATGTATCTCTGACAAAGAGTACGAGCA 519

Qy 361 gcaagtacagaagcagacacaaagacttctcgttttgcattgtgagatgagcatgagagga 420
Db 520 GCAAGTACAGAGCAGACACAAAGACTTCTCGTTTGTATATTGAGATGAGCATGAGAGGA 579

Qy 421 gatatattgagagaattgttcatttaccagaaacacctgtgactctggggaagataaacct 480
Db 580 GATATATTGAGAGATTGTTTCATTTTACAGAAACCTGTGATCTCTGGGAGAGATAAACCT 639

Qy 481 gagccagagacttggaagacttaataaataaggaagaaataagggctccatctt 540
Db 640 GAGCCAGACGATCTTGGAAAGTCAATTAATAATGGGAAAGAAATGGGCTCCATCTT 699
```

```

QY 541 cctgaacaagtacagaatgaatcaaatcaatgaagaaagaatgagtgagctatgtatt 600
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 CCTGAACAAGTACAGAATGAAATCAATCAATCAGTGAAGAACAGAGATGAGTGAGCTATGTATT 759

QY 601 gattttaacaaaactcaatcagagatgata 631
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 760 GA-TTTGACAAAAGACTCAATCAGGATGATA 789

RESULT 3
LOCUS BG180236 579 bp mRNA EST 06-FEB-2001
DEFINITION 602329846f1 NIH_MGC_91 Homo sapiens cDNA clone IMAGE:4431130 5',
      mRNA sequence.
ACCESSION BG180236
VERSION BG180236.1 GI:12686939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10185 row: 1 column: 11
High quality sequence stop: 579.
Location/Qualifiers
1..579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4431130"
/clone_lib="NIH_MGC_91"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: prostate; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.4 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 158 a 117 c 152 g 152 t
ORIGIN
source

Query Match 27.0%; Score 570; DB 11; Length 579;
Best Local Similarity 100.0%; Pred. No. 1.8e-129;
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1044 gggttttgaatgatgggaaatacaatgcctggagatctattactacatcagac 1103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 GGGTTTGAATATGATGGGAAATCAATGCCCTGGGATCTATATTACTACATGACTCAGAC 69

QY 1104 agaggaactcaagtattccatagacaaagagttcctcaagggaatactcccattgaggt 1163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 AGAGGAACCTCAAGTATTCATATAGACACAGAGTTCTCAAGGAATACTTCCCAATTGAGGT 129

QY 1164 ggtcactgaagccttgctgaacacaccaccagagtggttggaacttcaattgaaacaaat 1223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 GGTCACTGAAGCCTTGCTGAACACCTACCAGGAGTGTGTGGACTTTCATTTGAACAAT 189

QY 1224 gacagatgctcattgtttggaacaaagagtggttacactttacatctgtaaggataagctac 1283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 190 GACAGATGCTCATGTGTTGGAACAAGAGTGTACACTTTATACATTTTATACATGTAAGGATAAGCTAC 249

```

```

QY 1284 agagaagattgggacagcttctatttgagacctctatccaagggaagaaatacaatca 1343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 250 AGGAGAAGTATTGGGACAGTTCTATTGGACCTCTATCCAAAGGAAGAAATAFACAATCA 309

QY 1344 tgcggccttcctcggtctccagcctggctgcttctgctgatggaagccgagatgagc 1403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 TCGGGCCTGCTTCGGTCTCCAGCCTGGCTGGCTTCTGCTGATGGAAGCCGATGATGCC 369

QY 1404 agtgctgcctcctcggtgaaacttctcacagccagtgccagtgccctctctctcag 1463
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 AGTGCTGCCCTCGTGGTGAACCTTCTCACAGCCAGTGGCAGGCTGCTCCCTCTCTCTGAG 429

QY 1464 acacagcaggtgaggaacttacttctcagtgagtttggtcagtgatgcacagattgtgc 1523
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 430 ACACGACGAGGTGAGGACTTACTTTTCATGAGTTGTGCTACGTTGATGCATCAGATTTGTGC 489

QY 1524 acagactgatttgcagcagatttagcgggaacaaatggaagaaactgactttgtagagtgcc 1583
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 ACAGACTGATTTTGCACGATTTAGCGGAACAAATGTGGAACCTGACTTTGTAGAGTGCC 549

QY 1584 atcgcaaatgcttgaataattgggtgtggga 1613
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 550 ATCGCAAAATGCTTGAATAATTGGGTGTGGGA 579

RESULT 4
LOCUS BG666754 902 bp mRNA EST 30-APR-2001
DEFINITION DRABUG01 Rat DRG Library Rattus norvegicus cDNA clone DRABUG01 5',
      mRNA sequence.
ACCESSION BG666754
VERSION BG666754.1 GI:13888661
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 902)
Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,
Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and
Zhang,X.
Distinct gene expression profiles of rat dorsal root ganglion
induced by peripheral nerve axotomy
Unpublished (2001)
Contact: Zhang Xu
Laboratory Of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
Location/Qualifiers
1..902
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRABUG01"
/clone_lib="Rat DRG Library"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
BASE COUNT 214 a 238 c 239 g 211 t
ORIGIN

```





/note="Organ: testis; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

203 a 166 c 210 g 166 t

BASE COUNT  
ORIGIN

```
Query Match      22.3%; Score 472.6; DB 11; Length 745;
Best Local Similarity 96.3%; Pred. No. 1.6e-105;
Matches 548; Conservative 0; Mismatches 14; Indels 7; Gaps 6;

Qy 1 atgacgcgcgcgtgcttggctgtgcgaagcctccgcagagtggtggtccaggatt 60
    |||
Db 153 ATGATCGCCCGTGCCTTTGGG-TGTGCGAAGCCTCCGCAGAGTTGTTGGTTCCAGGATT 211
    |||

Qy 61 ttactcgaatgcagtttaggaagaagtgatgtctctctcttcaggcaatgtctctat 120
    |||
Db 212 TTACTCAGAATGACGTTAGGAAGAAGATGATGCTCTCTCTTCAGGCAATGCTTCCTAT 271
    |||

Qy 121 actgtgctgcagaaatgttttaagatggggtctttcaccagagcaaatataaacaaga 180
    |||
Db 272 ACTGTGGCTGCAGAAATGTTTAAAGATGGGATCTTTCACCAGAGCAAAATTAACAACAGA 331
    |||

Qy 181 actgaggagctcattgtgcagaccacaaacagtgctacgatgctgttggaaatgctcggtatt 240
    |||
Db 332 ACTGAGGAGCTCATTGTGCAGACCAACACAGCTGTACGATGCTGTGGAAATGCTCGGTATT 391
    |||

Qy 241 gaggaaatcattcagagaactgtctgcaggaactgtgcagatgtagaagtaaatatata 300
    |||
Db 392 GAGGAAGTAACTTACGAGAACTGTCTGCAGGACCTGGCAGATGTAGAAATTAAGTATATA 451
    |||

Qy 301 gtgaaagaccatgtacgtctccacagcatgtatctctgcagaaagaaagtaacagca 360
    |||
Db 452 GTGGAAGGACCATGCTAGACTTTCCCGAGCATGTATCTCTGCAAGAAGTAGACGAGCA 511
    |||

Qy 361 gcaagtacagaagcagacaaaagacttctcgttttggattattgagatgagcatgagagga 420
    |||
Db 512 GCAAGTACAGAAGCAGACAAAAGACTTCTCGTTTGTATATTGATGAGCATGAGAGGA 571
    |||

Qy 421 gata-tattgagagaattgttcattt--acagaaacactgtatctgtgggaagataaaa 477
    |||
Db 572 GATATTATTTTTCAGAGAATGTTTTCATTTTACAGGAAACCTGTGATCTGGGGAAGATAAAA 631
    |||

Qy 478 cctgagccacacatacttgaaaagtcaattaaaatggggaagaaatgggctccat 537
    |||
Db 632 ACTGAGG-CACACGATAC-TGGGAAGTCAATTACAAATGGGGCAAGAAATGGCGTCCAT 689
    |||

Qy 538 ctctcgaacagtcagaaatgaatcaaa 566
    |||
Db 690 CTT-CTGAACATGTCAGAAATGAATCAA 717
    |||
```

```
RESULT 8
BE311015 559 bp mRNA EST 26-OCT-2000
LOCUS 601088224F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3483124 5',
DEFINITION mRNA sequence.
ACCESSION BE311015
VERSION BE311015.1 GI:9171599
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 559)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
```

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM8513 row: p column: 05  
High quality sequence stop: 557.

FEATURES  
source

```
1..559
Location/Qualifiers
/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone_image="3483124"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
```

BASE COUNT 139 a 136 c 146 g 138 t  
ORIGIN

Query Match 21.5%; Score 455.6; DB 10; Length 559;  
Best Local Similarity 88.5%; Pred. No. 2.2e-101;  
Matches 494; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy 1167 cactgaaggctgtgcgaacactaccagagtggttgaggactttcattgaacaaatgac 1226
    |||
Db 1 CACTGAAGGCCCTTAGACATCTACAGGAGCTGTGGGACTTTTCATTGAACAAGTGGC 60
    |||

Qy 1227 agatgctcatgttggaaacagaggtttacactttactgtgaaggataaagctacagg 1286
    |||
Db 61 TGATGCCCATGTTTGGAAATAAGAGTGTTCACCTTACCCGTGAAGGATAAAGTACTTGG 120
    |||

Qy 1287 agaagtattggagcagttctatttggacctctatccaaagggaagaaatacaaatcagc 1346
    |||
Db 121 AGAAGTGTGGGGCAGTTCTACCTGGACCTCTATCCAAGGGAAGGAAATACAACATGC 180
    |||

Qy 1347 ggcctgctcgtgtctccagcctgctcctctcctgatgaaagccgagatgagcagt 1406
    |||
Db 181 GGCCTGCTTTGGTCTCCAGCCAGGCTGCCCTTCTCCCTGATGGCAGTCGGATGATGCTGT 240
    |||

Qy 1407 ggcctgctcgtgtgaaacttctcacagcagtgccaggtcgtcctctcctctgagaca 1466
    |||
Db 241 GGCCTGCCCTGGTGCACACTTCTCTACCCCATAGCAGCGCGGCCCTCTCTCCTCAGACA 300
    |||

Qy 1467 cgacgagtgaggacttactttcatagtttgggtcacgtgatgcacagatttggcaca 1526
    |||
Db 301 CGACGAAGTGCGAACCTTACTTCCATGAGTTGGTGCATGTCATCAGATCTGTGCACA 360
    |||

Qy 1527 gactgattttgcacgatttagcgaaacaaatgtggaactgactttgtagaggtccatc 1586
    |||
Db 361 GACTGACCTTTCACGATTTAGTGGAAACAAATGTGGAATGCTTTGTAGAGTGCCATC 420
    |||

Qy 1587 gcaaatgttgaataatgggtgtggagcgtcgattccctccgaaagatgtgcaaacatta 1646
    |||
Db 421 ACAATGCTTGAATACTGGGTGTGGGACATTTGACTCCCTGCGAAACATGTCNAACATTA 480
    |||

Qy 1647 taaagatggaagccctatttgacagacgactctgttgaataacttgttcttctagctggt 1706
    |||
Db 481 TAGAGACCGACACCTTATCAGAGAGAGCTGCTGGAGAAGCTTGTGGCTTCGAGACTGT 540
    |||

Qy 1707 caacacaggtctctgac 1724
    |||
Db 541 CAACACAGGTCTCTCTGAC 558
    |||
```

RESULT 9

AW786845  
LOCUS 120431 MARC 1P1G Sus scrofa cDNA 5', mRNA EST 09-JUL-2000  
DEFINITION AW786845  
ACCESSION AW786845  
VERSION  
KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
1 (bases 1 to 509)  
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,  
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.  
and Keele,J.W.  
TITLE Design and use of two pooled tissue normalized cDNA libraries for  
EST discovery in swine  
JOURNAL Unpublished (2000)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCCAGTCAGCAG  
Plate: 39 row: N column: 18  
Seq primer: ATTAGTGACACTATAG.  
FEATURES  
Location/Qualifiers  
1..509  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone\_lib="MARC 1P1G"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 11, 13, 15, 20,  
and 30 embryos."  
BASE COUNT 155 a 99 c 115 g 140 t  
ORIGIN  
source  
Query Match 20.4%; Score 431.4; DB 10; Length 509;  
Best Local Similarity 91.7%; Pred. No. 28-95;  
Matches 467; Conservative 0; Mismatches 41; Indels 1; Gaps 1;  
QY 1515 gattgtgcacagctga-tttgtcacgatttagcggaacaatgtgaaactgactttg 1573  
DB 1 GATTGTGCGCAGACAGCTTCTGCGAGATTAAATGACACACATGTGGAAACAGACTTTG 60  
QY 1574 tagagtgccatgcgaatgcttgaataattgggtgtgggacgtcgattccctccgaagat 1633  
DB 61 TAGAGGTGCGCATCACAAATGCTTGAACACTGGGTGTGGGACACTGATTCCTCCGAAGGC 120  
QY 1634 tgcacaaacattataagatggaagccctattgcagacgatctgcttgaaaaactgttg 1693  
DB 121 TGTCAAAACATTATAAAGATGGAAGCCCTATTACAGATGATCTGCTGAAAAACTTGTG 180  
QY 1694 ctctaggtcgttcaacacaggtcttctgacctgcgcagattgttttgacaaagtig 1753  
DB 181 CTCTAGACTGGTCAACACAGAGTCTCTGACCTTAGCCAGATGTGTTTGTAGCAAAAGTTG 240  
QY 1754 atcagctctctacacacacacacacgcgtgagatgctgcaagtgaatgcaaaactgct 1813  
DB 241 ATCACCTCTCCACACCAACACTTCGCTGGATGCTGCAAGTGAATATGCCAATACTGCA 300  
QY 1814 cagaaatattagaggtgcagctactccaggcagcaaatatgccagctacacttggacatt 1873  
DB 301 CAGAAATTTTAGGTGTGGAACCTACTTCCAAGCACAATAATGCCAGCTACTTTTGGGACT 360

QY 1874 tggcagggggatcagtgcccaattatttgatattctttggagtgaagtattttccatg 1933  
DB 361 AGGCAGGGGGATGATGATGCCATATATATGATATGTTGGAGTGAAGTCTTTTCCATGG 420  
QY 1934 atatgttttacagctgttttttaaaagagggaataatccacagaggttggaatgaaat 1993  
DB 421 ACATGTGATATAACTCTTTTAAAAAAGATGGATAATGAATCCTCGAGTTGGAGTGAATGAAT 480  
QY 1994 acagaaacctaatctgaacctgtggga 2022  
DB 481 ACTGAAACCTAATCTCTGAAACCTGGGGGA 509  
RESULT 10  
AW732454  
LOCUS 570 bp mRNA EST 21-APR-2000  
DEFINITION bb06c03.y1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:2958628 5'  
similar to SW:NEUL\_RABIT P42675 NEUROLYSIN PRECURSOR ;, mRNA  
sequence.  
ACCESSION AW732454  
VERSION AW732454.1 GI:7632786  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 570)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
image.lnl.gov/image/html/iresources.shtml  
Seq primer: -40RP from Gibco  
High quality sequence stop: 393.  
FEATURES  
Location/Qualifiers  
1..570  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:2958628"  
/clone\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 130 a 138 c 161 g 140 t  
ORIGIN  
source  
Query Match 19.8%; Score 419.8; DB 10; Length 570;  
Best Local Similarity 94.0%; Pred. No. 1.4e-92;  
Matches 436; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 1 atgatcgcccggtgccttttggctgtgcgaagcctccgagagttggtgtccaggatt 60  
DB 106 ATGATCGCCCGGTGCCTTTTGGCTGTGCGAAGCCTCGCAGAGTTGGTGTCTCCAGGATT 165  
QY 61 ttactcagaatgacgttaggaagaagatgagtctctctcttcaggcaaatgtcttctat 120  
DB 166 TTACTCAGAATGACGTTAGGAGAGAAGTATGATGTCCTCTTCAGGCAATGTCTCTCTAT 225



Qy 121 actgtggcgtgcagaaatgttttaagatggtgatctttccaccagagcaaatataaacaaga 180  
Dy 226 ACTGTGGCTGGCAGAAAATGTTTTAAGATGGGATCTTTCACCAGAGCAAAATTTAAACAGA 285  
Qy 181 actgaggagctcattgtgcaaacacacaggtgtacgatgtgttggaatctcggtatt 240  
Dy 286 ACTGAGGAGCTTCATGTGTGCACACCAACAGGTGTACGATGCTGTGGAAATGCTCGGTATT 345  
Qy 241 gaggaagttaacttacgagaactgtctgcagagcactggcagatgtagaagttaagtatatata 300  
Dy 346 GAGGAAGTAACTTACGAGAAGTGTCTGCAGGCACTGGCAGATGTGAAGTTAAGTATTTA 405  
Qy 301 gtgaaaggaccatgctagactttcccccagcatgtatctctgcacaaagaagtcagagca 360  
Dy 406 GTGTGAAGGACCATGCTAGACTTTCCCCAGCATGTTTCTCTTGCACAGAGGTTCGAGCC 465  
Qy 361 gcaagtacagagcagacacaaagaactttctctgttttgatattgatagagcatgagagga 420  
Dy 466 GCANGGTCAGGAAGCAGACAAAAGACTTCTCGTGTGTGATATTGGGATGGGCTTGGAGGG 525  
Qy 421 gatattattgagaaattgttcattacagagaaacctgtgatct 464  
Dy 526 GATATTTTGTAGAGAAATGGTCACTTTTCAGGAACCCCTGGATCT 569  
  
RESULT 11  
AI386253  
LOCUS AI386253 511 bp mRNA EST 27-JAN-1999  
DEFINITION mm42g01.y1 Stratagene mouse melanoma (#937312) Mus musculus cDNA  
clone IMAGE:524208 5' similar to SW:NEUL\_RAT P42676 NEUROLISIN  
PRECURSOR ; mRNA sequence.  
ACCESSION AI386253  
VERSION AI386253.1 GI:4199716  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 511)  
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person  
B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter  
E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:318036  
This read is a RESEQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Putative full length read  
vector to vector length is 741  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Gibco  
High quality sequence stop: 351.  
FEATURES  
source Location/Qualifiers  
1..511  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:524208"  
/clone.lib="Stratagene mouse melanoma (#937312)"  
/tissue\_type="melanoma"  
/dev\_stage="M2 cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="organ: skin; Vector: pBluescript SK-; Site\_1: EcoRI

; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT. From M2 cells, a highly metastatic derivative of the  
K-1735 (mouse) melanoma. Average insert size: 1.0 kb;  
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG  
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."  
BASE COUNT 140 a 115 c 134 g 122 t  
ORIGIN  
  
Query Match 19.7%; Score 416.8; DB 10; Length 511;  
Best Local Similarity 88.4%; Pred. No. 7.6e-92;  
Matches 451; Conservative 1; Mismatches 58; Indels 0; Gaps 0;  
  
Qy 1603 tgggtgtgggacgtcgattccctccgaagattgtcaaaacattataaagatggaagccct 1662  
Dy 2 TGGGTGTGGGACATTGACTCCTCGGAAACATGTCAAAACATTATAGAGAGGGACACCCCT 61  
Qy 1663 attcagacgacgtcgcttgaataaactgttctctagctggtggtcaacacaggtctcttg 1722  
Dy 62 ATCACACGACGAGCTGTGGGAAAGCTTGTGGCTTCGAGACTGGTCAACACAGGTTCTCTG 121  
Qy 1723 acctgcgcagagattgtttgagcaaaattgatcagctctcttcataccaacacacatcgctg 1782  
Dy 122 ACGCTTCGCCAAATGTTTTCAGCAAAAGTTGACCACTCCCTCCATACCAATGCTGCGCTG 181  
Qy 1783 gatctgcaagtgaatgccaatactgctcagaataattaggagtgctgagctactcca 1842  
Dy 182 GATCGCGGAGCAATACGCTAAATACTGTCACAGAAATTTTGGGAGTTGCGAGCTACGCCA 241  
Qy 1843 ggcacaaatagccagctacaccttggacatttgcagggggagatcgatggccaattattat 1902  
Dy 242 GGCACAAATATGCCAGCTTACCTTTGGACATTTTGGCAGGAGGTATGACGCCCATTTAT 301  
Qy 1903 ggatatcttggagtgaagtatttccatgatgttttacagctgttttaaaaaaagaa 1962  
Dy 302 GGATATCTTTGGAGTGAAGTCTTTTCCATGGACATGTTTCACAGCTGTTTCAGAAAAGAA 361  
Qy 1963 gggataatgaatccagaggttggaaatgaatacagaacacctaactctgaaacctgggggga 2022  
Dy 362 GGGATTATGAATCAGAGGTTTGGAAATGAATACAGAAACCTAACTCTGAAGCCTGGGGGG 421  
Qy 2023 tctctgacgagcagtcagacatgctccacaatttcttgaacgtgagccaaacacaaagcgc 2082  
Dy 422 TCCCTGGACGCATGGACATGCTCCAGAAATTTCTTGGCAACGTGAGCCAAACACAGAAAGCA 481  
Qy 2083 ttctaattgagtagagcgtgcagtctgctgcg 2112  
Dy 482 TTCTGTGATGATCGAGGCTGAATGCTTCG 511  
  
RESULT 12  
AW003054/c  
LOCUS AW003054 757 bp mRNA EST 08-MAR-2000  
DEFINITION wg61h05.x1 NCI\_CGAP\_GC6 Homo sapiens cDNA clone IMAGE:2475801 3'  
similar to SW:NEUL\_PIG Q02038 NEUROLISIN PRECURSOR ; mRNA  
sequence.  
ACCESSION AW003054  
VERSION AW003054.1 GI:5849892  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 757)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima





```

Qy 1902 tggatatcttggagtgagtgattttccatcgatattgttttacagctgtgttttaaaaaaaga 1961
Db 440 TGGATATCTTTGGAGTGAAGTGTTCATCGACATGTTTCACAGCTGTTTAAAAAAGA 381
Qy 1962 agggataatgaatccagaggttggaatgaataacagaaacacctaactcctgaaacctggggg 2021
Db 380 GGGGATCATGAATCCAGAGGTGGAATGAATGAATACAGAAACTTAATCCTGAAGCCTGGGGG 321
Qy 2022 atctctgacggcatggacatgctcccaattcttgaacgtgagccaaacccaaaaagc 2081
Db 320 GTCCTGGACGGCATGGACATGCTCCAGAAATTTCTTGCACGTGAGCCAAACCAAAAGC 261
Qy 2082 gttcctaagttagtaggcctgcgtctgctgta 2115
Db 260 ATTCTAATGAGTCGAGCCCTGAATGGTTCGTAA 227

RESULT 14
BF719641
LOCUS BF719641 675 bp mRNA EST 03-JAN-2001
DEFINITION mab42f07.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone
IMAGE:3973020 5' similar to SW:NEUL_RAT P42676 NEUROLYSIN PRECURSOR
;A mRNA sequence.
ACCESSION BF719641 GI:12020643
VERSION BF719641
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 675)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

MGI:1473052
Seq primer: -40RP from Gibco
High quality sequence stop: 457.
Location/Qualifiers
1 . 675
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3973020"
/clone.lib="Soares_NMEBA_branchial_arch"
/tissue_type="branchial arches"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/notes="vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCATGATTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 186 a 161 c 187 g 137 t 4 others
ORIGIN

Query Match 17.68; Score 373.2; DB 11; Length 675;
Best Local Similarity 77.3%; Pred. No. 4.1e-81;
Matches 450; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

```

```

Qy 1 atgatcgcccggtgccttttggctgtgcgaagcctcccgagagttgggtgggtccaggatt 60
Db 94 ATGATCACCCCTGTGCTTTCGGCTCTGCGAGGCCTCCACAGAGCTGGTGGTTCAGGATT 153
Qy 61 ttactcagaatgaagtttaggaagaagtgatgtctctctcttcaggcaaatctctccat 120
Db 154 CGGTCTAGAATGAGCTTTGGGAGGAGGCGGCATCTCCTCTTCAGGCCATGCTCTTCCTAC 213
Qy 121 actgtgctgcgaataatgttttaagatgggatctttcaccagagcaaatataaacaaga 180
Db 214 ACCCGCGCAGCAGGAATGTTCTCAGATGGGACCTTTCTCCAGAGCAGATCAGGACGAGG 273
Qy 181 actgagagctcattgtgcagaccacacaggtgtacgatgctgtgtgggaatcctcggtatt 240
Db 274 ACAGAGGAGCTCATAGCACACACCAAGCAGGTGTATGATACCGTGGGACCAATCAACCTG 333
Qy 241 gaggaagtaacttacgagaactgtctgcaggcaactggcagatgtagaagtagaagtataata 300
Db 334 GAGGATGTAACCTACGAGAACTGTCTGAGGTGGTGGCTGACATAGAAGTGAAGTACATA 393
Qy 301 gtgaaaggacactgctagactttcccccagcatatcctctgcacaaagaagtagcagca 360
Db 394 GTGGAAGGACCATGCTGGACTTCCCTCAGCACGCTGCTCTGACAGAGAAGTGGCGGCT 453
Qy 361 gcaagtacagaagcagacaaagacttctctgtttttgtattgtatgtgagatgagcatgagga 420
Db 454 CCAAGCAGAGGCGGACAAAAGGCTGCTCTGTTGTATATTGAGATGAGCATGAGAGAA 513
Qy 421 gatatttgagagaattgttcatttacgagaacctgtgatctggtgggaagataaaacct 480
Db 514 GATGATTTTCAGAAAATAGGTGATGTACAGAAACGTCGAATTTGGAGAAGATAAAGCT 573
Qy 481 gagccagacatacttgaaaagtcaattaaaatggggaagaaatgggctccatctt 540
Db 574 GAGCCAGGCGCATCTTTGGAGAAATCAATTAANATGTTAAAGATGGAGTGGACCTCATTT 633
Qy 541 cctgaacaagtacagaatgaaatcaatcaatcaatgaagaaga 582
Db 634 CAACAACNTTATAANATGAATTNAATAAATGAATAGAGAAA 675

RESULT 15
AA152501
LOCUS AA152501 417 bp mRNA EST 01-AUG-1997
DEFINITION z007a04.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA
Clone IMAGE:566958 5' similar to SW:NEUL_RABIT P42675 NEUROLYSIN
PRECURSOR ;, mRNA sequence.
ACCESSION AA152501
VERSION AA152501 GI:1718695
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 417)
REFERENCE
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., DuBoque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Rohlfing,T.,
Woodward,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
TITLE
JOURNAL
MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

Search completed: January 14, 2002, 19:34:07  
Job time: 2176 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2002, 19:34:11 ; Search time 380.46 seconds

(without alignments)  
10766.075 Million cell updates/sec

Title: US-09-833-782-1

Perfect score: 2115

Sequence: 1 atgatgcccggtgcctttt.....gagccctgcctgctcgtga 2115

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1481224 seqs, 968335871 residues

Total number of hits satisfying chosen parameters: 2962448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*

1: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US06\_NEW\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US10\_NEW\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	371.8	17.6	408	6	US-09-933-524A-18292	Sequence 18292, A
2	215.4	10.2	222	6	US-09-826-734-43	Sequence 43, Appl
3	206	9.7	543	6	US-09-898-888A-27547	Sequence 27547, A
4	190.2	9.0	449	6	US-09-933-524A-11054	Sequence 11054, A
5	176.8	8.4	479	5	US-09-939-397-39435	Sequence 39435, A
6	126	6.0	471	5	US-09-939-397-39669	Sequence 39669, A
7	121.8	5.8	468	6	US-09-933-524A-7235	Sequence 7235, Ap
8	98.6	4.7	511	6	US-09-388-906A-20678	Sequence 20678, A
9	96.2	4.5	391	6	US-09-933-524A-3065	Sequence 3065, Ap
10	94.8	4.5	388	6	US-09-898-888A-37608	Sequence 37608, A
11	87.8	4.2	270	5	US-09-927-875A-3719	Sequence 3719, Ap
12	86.8	4.1	269	6	US-09-839-976A-149	Sequence 149, App
13	79.8	3.8	2513	5	US-09-865-439A-20408	Sequence 20408, A
14	78.2	3.7	637	5	US-09-865-439A-67646	Sequence 67646, A
15	77.4	3.7	9574	6	US-09-897-516-2236	Sequence 2236, Ap
16	77.4	3.7	9574	6	US-09-897-516-2237	Sequence 2237, Ap
17	77.4	3.7	9574	6	US-09-897-516-2238	Sequence 2238, Ap
18	77.4	3.7	9574	6	US-09-897-516-2239	Sequence 2239, Ap
19	77.4	3.7	9574	6	US-09-897-516-2240	Sequence 2240, Ap
20	77.4	3.7	9574	6	US-09-897-516-2241	Sequence 2241, Ap
21	77.4	3.7	9574	6	US-09-897-516-2242	Sequence 2242, Ap
22	77.4	3.7	9574	6	US-09-897-516-2243	Sequence 2243, Ap
23	77.4	3.7	9574	6	US-09-897-516-2243	Sequence 2243, Ap
24	75	3.5	2037	6	US-09-806-866-56	Sequence 56, Appl
25	73.4	3.5	1431	6	US-09-806-866-50	Sequence 50, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-933-524A-18292  
; Sequence 18292, Application US/09933524A  
; GENERAL INFORMATION:  
; APPLICANT: Drmanac, Radoje T.  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Dickson, Mark  
; APPLICANT: Jones, Lee W.  
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained  
; TITLE OF INVENTION: From Various Libraries  
; FILE REFERENCE: 774  
; CURRENT APPLICATION NUMBER: US/09/933,524A  
; CURRENT FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 09/528,409  
; PRIOR FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 116231  
; SOFTWARE: Hy-patent.pl Version 3.1  
; SEQ ID NO 18292  
; LENGTH: 408  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-933-524A-18292

Query Match 17.6%; Score 371.8; DB 6; Length 408;  
Best Local Similarity 98.9%; Pred. No. 8.4e-90;  
Matches 373; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 709 gacaagtataaaattacctttaaataatccacactatttccctctcatgaagaattgtgt 768  
Db 32 gacaagtataaaattacctttaaataatccacactatttccctctcatgaagaattgtgt 91  
QY 769 atccctgaaccagaagaaggatgaaatggcttttaatacaaggtgcaagaggaaaac 828  
Db 92 atccctgaaccagaagaaggatgaaatggcttttaatacaaggtgcaagaggaaaac 151  
QY 829 accataattttcagcagctactccactgcgaacaaagtgccaaactactcgtttat 888  
Db 152 accataattttcagcagctactccactgcgaacaaagtgccaaactactcgtttat 211  
QY 889 agcacacatgctgacttccttgaaatgaacactgcaagagcacaagcgcgtaaca 948  
Db 212 agcacacatgctgacttccttgaaatgaacactgcaagagcacaagcgcgtaaca 271  
QY 949 gcytttctagatgatttaagccaggaagttaaacccttgggtgaagcagaacagagttt 1008  
Db 272 gcctttctagatgatttaagccaggaagttaaacccttgggtgaagcagaacagagttt 331

Sequence 54, Appl  
Sequence 60, Appl  
Sequence 423, Appl  
Sequence 3852, Appl  
Sequence 3853, Appl  
Sequence 3854, Appl  
Sequence 3855, Appl  
Sequence 3856, Appl  
Sequence 3857, Appl  
Sequence 3858, Appl  
Sequence 3859, Appl  
Sequence 3860, Appl  
Sequence 3861, Appl  
Sequence 3862, Appl  
Sequence 3863, Appl  
Sequence 3864, Appl  
Sequence 3865, Appl  
Sequence 3866, Appl  
Sequence 3867, Appl  
Sequence 3868, Appl





Db 150 cccaggaaacacgctgctgcaacctcgccatcgccatgaggtggctacacgcccagt 209  
QY 1898 attatgatatcttggagtgaaagtattttccatggtatgttttacagctgttttaaaa 1957  
Db 210 actacgggtacctgtggagcaggtgtattccatggacatgttccacacgcgttcaagc 269  
QY 1958 aagaaggataatgccacccagaggttgaatgaaatcacagaaacctaactctgaaacctg 2017  
Db 270 aggagggcgccctgaaacagcaaggttggcatggtattacagaagctgcatactgagacccg 329  
QY 2018 ggggactctggacggcatggacatgctccacaatttcttgaacgtgagcccaacaaa 2077  
Db 330 gcggttccggagtcgacgcgcctatgctggcgcttcttccggcgctgaccccaagcagg 389  
QY 2078 aagcgttccctaatgagtagaggcctgcagt 2107  
Db 390 acgcttctctctgagcaaggggctgcagg 419

## RESULT 7

US-09-933-524A-7235

; Sequence 7235, Application US/09933524A

; GENERAL INFORMATION:

; APPLICANT: Drmanac, Radoje T.

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Crain, Birgit

; APPLICANT: Dickson, Mark

; APPLICANT: Jones, Lee W.

; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained

; TITLE OF INVENTION: From Various Libraries

; FILE REFERENCE: 774

; CURRENT APPLICATION NUMBER: US/09/933,524A

; CURRENT FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 09/528,409

; PRIOR FILING DATE: 2000-05-17

; NUMBER OF SEQ ID NOS: 116231

; SOFTWARE: Hy-patent.pl Version 3.1

; SEQ ID NO 7235

; LENGTH: 468

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(468)

; OTHER INFORMATION: n = A,T,C or G

US-09-933-524A-7235

Query Match 5.8%; Score 121.8; DB 6; Length 468;  
Best Local Similarity 64.2%; Pred. No. 1.7e-22;  
Matches 183; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 128 ctggcagaaatgttttaagatggagctttccacagagcaaaattaaacagaactgagg 187  
Db 184 ctgtgttaacagacctgcgtgggacacctgagtgccacagacagatagagagcgccagg 243  
QY 188 agctcatgtgcagacacaaacagggtgtacgactgctgttggaatgctgctgtattgaggag 247  
Db 244 agctcatgcagacacaaacaggcgctgtatgaccagggttggcaccaggagtttgaggagc 303  
QY 248 taacttagaagaactgtctgcagccactggcagatgtgagaagtaagtatatagtggaag 307  
Db 304 tgtctcagagacagcgtcaaggcgctggccgagtggtgagggtccactacacagttcaga 363  
QY 308 ggaacatgctagactttcccacagatgtatctctgcacaaagaagtagacagacgaagta 367  
Db 364 ggaatctctgacttcccacagcatgtttcccctcccaaggacacatccggacagccaga 423  
QY 368 cagaagcagacaaagactttctgttttatagatgagca 412  
Db 424 cagagggccgacaagaagctctctgagttcgcagctggagatgagca 468

## RESULT 8

US-09-388-906A-20678

; Sequence 20678, Application US/09388906A

; GENERAL INFORMATION:

; APPLICANT: Havukkala, Ilkka

; APPLICANT: Shenk, Michael Andrew

; TITLE OF INVENTION: Polynucleotides Isolated From Plants and

; TITLE OF INVENTION: Methods For Their Use

; FILE REFERENCE: 11000.1013U

; CURRENT APPLICATION NUMBER: US/09/388,906A

; CURRENT FILING DATE: 1999-09-01

; NUMBER OF SEQ ID NOS: 24843

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 20678

; LENGTH: 511

; TYPE: DNA

; ORGANISM: Eucalyptus grandis

US-09-388-906A-20678

Query Match 4.7%; Score 98.6; DB 6; Length 511;  
Best Local Similarity 52.0%; Pred. No. 3e-16;  
Matches 244; Conservative 0; Mismatches 224; Indels 1; Gaps 1;

QY 1077 ggatctattactacatgactcagacagagggaaactcaagtattccatagaccagagtt 1136  
Db 35 ggatctattgtattattaaagacatttcaagagcagcactttgatagtattttggagc 94  
QY 1137 cctcaagaatacttcccaattgaggtggtcactgaaggcttgcgaacctaccagga 1196  
Db 95 cattaagcaatatttcccggtcaaatagttctctatcagggcatctttaaattttccaga 154  
QY 1197 gtgttgggactttcatttgaacaaa-tgacagatgctcatgttttggaaacagagtta 1255  
Db 155 cctatttgggttgagatttgaggaaattgccgacctgaggtcgtggttgatgttc 214  
QY 1256 cactttactgtgaaggataaaagctacagaggaagatttgggacagttctatttggacc 1315  
Db 215 gattatttccagttattgacttaagtccggtgactcctcctggggcatttctaccttata 274  
QY 1316 tctatcaagggaagaaatacaatcatcgctcgtcgtctccagctggtcgcc 1375  
Db 275 tgtatagaagaagaaagacggtcacacatcggtggtgctcttcacagacagatcat 334  
QY 1376 ttctgcctgatggaaagccggatgatggcagtggtgctcctcgtggtgaaacttccacgc 1435  
Db 335 tatcttttaacaatgttggccagataccagtggtatgtctaataactcaattcccaagg 394  
QY 1436 cagtggcaggtcgtccctctctctctcagacagcagcagaggtgaggaacttactatgag 1495  
Db 395 atatcagtggtcactcctggttctctacgattctccgaaagtagtcaacttttccatgagt 454  
QY 1496 ttgtcagctgatcacatcagatttggcacagactgattttgcagatt 1544  
Db 455 ttggacatgtgatgcagcatatctgcacaagggcatcatatttggcagatt 503

## RESULT 9

US-09-933-524A-3065

; Sequence 3065, Application US/09933524A

; GENERAL INFORMATION:

; APPLICANT: Drmanac, Radoje T.

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Crain, Birgit

; APPLICANT: Dickson, Mark

; APPLICANT: Jones, Lee W.

; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained

; TITLE OF INVENTION: From Various Libraries

; FILE REFERENCE: 774

; CURRENT APPLICATION NUMBER: US/09/933,524A

; CURRENT FILING DATE: 2001-08-20

; PRIOR APPLICATION NUMBER: 09/528,409







; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LTB3587-284-Q1-K6-C10  
US-09-865-439A-67646

```
Query Match          3.7%; Score 78.2; DB 5; Length 637;
Best Local Similarity 52.2%; Pred. No. 1e-10;
Matches 198; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

QY 1396 atgatggcagtgagcgcctcgtgtggaacttctcacagccagtggcaggtcgccctct 1455
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 aggcgcctgttgcccatatagtgtgcaatcagactccaccagttggcgagaagccagc 243
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1456 ctccctgagacacgacgaggtgaggacttactttcatgagtttggtcacgtgatgcacag 1515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 244 ctcatgaccttcggtgaggttgaaactgtgtccatgaatttgggcacgccccttcagcac 303
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1516 atttgtcacagactgattttgcacgatttagcggaacaaat---gtgaaactgacttt 1572
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 atgcttactaaacaagatgaaggctttgtgtggttcgttcgttcgttcgttcgttcgttc 363
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1573 gttagaggtgccatcgcaaatgttgaaattgggtgtgggacgtcgattccctccgaaga 1632
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 364 gttagaggtaccctccagttcatggagaactggtgctatcacagaataactcttttgagc 423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1633 ttgtcaaaacattataagatggaagccctattgcagacgatctgttgaaaaacttgtt 1692
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 424 attgcaaaacattatgaaccggtgaaacccctccagagaaatttatcggaagcttgta 483
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1693 gcttctaggtggtcaacacaggttcttgcacctgcgcagattgttttgagcaaatgt 1752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 484 gctgcaaaagaatttcggtggtgaccttgagccctgagcctgcgtcagatacagatttgc 543
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1753 gatcagttcttcatacca 1771
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 544 gatagggagcttcatacaa 562
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: January 14, 2002, 21:26:15  
Job time: 6724 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 14, 2002, 18:43:21 ; Search time 18.52 Seconds  
(without alignments)  
855.417 Million cell updates/sec

Title: US-09-833-782-2

Perfect score: 3668

Sequence: 1 MIARCLLAVSRRLRVGSGRI.....FLKREPNOKAFRLMSRGLHAP 704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCrUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2979	81.2	709	5	PCT-US92-00731-13
2	2331	63.5	689	1	US-07-766-351-5
3	2331	63.5	689	1	US-08-059-032-5
4	2331	63.5	689	5	PCT-US91-07290-5
5	122	3.3	3248	1	US-08-353-700-1
6	122	3.3	3248	5	PCT-US95-16216-1
7	120	3.3	1388	2	US-08-685-576-1
8	116	3.2	892	1	US-07-977-434-12
9	116	3.2	892	1	US-08-458-819-12
10	116	3.2	892	5	PCT-US91-07035-12
11	115	3.1	477	4	US-09-128-142-2
12	115	3.1	505	4	US-09-128-142-4
13	115	3.1	872	1	US-08-766-014-2
14	113.5	3.1	643	2	US-08-245-511-47
15	113.5	3.1	643	2	US-08-600-993A-47
16	113	3.1	976	4	US-09-104-324B-4
17	113	3.1	1388	2	US-08-685-576-4
18	112.5	3.1	615	4	US-08-989-299-11
19	112	3.1	1164	4	US-08-923-992A-2
20	112	3.1	2285	4	US-09-308-375-2
21	112	3.1	2496	4	US-09-125-028-2
22	111.5	3.0	712	2	US-08-468-576B-17
23	111.5	3.0	712	2	US-08-468-579B-17
24	111.5	3.0	712	3	US-08-468-577B-17
25	111.5	3.0	893	2	US-08-706-702-3
26	111.5	3.0	893	3	US-08-706-706-3
27	111.5	3.0	1713	3	US-08-600-982-24

28	111.5	3.0	1713	5	PCT-US94-10261A-24	Sequence 24, Appl
29	111	3.0	1354	3	US-08-685-871-2	Sequence 2, Appl
30	111	3.0	1786	4	US-08-973-462-8	Sequence 8, Appl
31	110.5	3.0	422	4	US-08-646-693-2	Sequence 2, Appl
32	110.5	3.0	422	5	PCT-US96-06053-2	Sequence 2, Appl
33	110	3.0	351	1	US-08-402-217A-2	Sequence 2, Appl
34	110	3.0	351	1	US-08-700-178-2	Sequence 2, Appl
35	110	3.0	351	3	US-08-995-654-2	Sequence 2, Appl
36	110	3.0	2958	4	US-08-894-344C-2	Sequence 2, Appl
37	109.5	3.0	476	4	US-08-134-557D-2	Sequence 2, Appl
38	109.5	3.0	788	2	US-08-907-166-6	Sequence 6, Appl
39	109	3.0	475	2	US-08-484-200-2	Sequence 2, Appl
40	109	3.0	475	3	US-08-465-375-2	Sequence 2, Appl
41	109	3.0	1128	4	US-08-923-992A-6	Sequence 6, Appl
42	107.5	2.9	1264	1	US-07-789-915A-6	Sequence 6, Appl
43	107.5	2.9	1264	1	US-08-005-002C-6	Sequence 6, Appl
44	107.5	2.9	1264	1	US-08-487-203A-6	Sequence 6, Appl
45	107	2.9	1698	4	US-09-315-793-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1  
PCT-US92-00731-13  
; Sequence 13, Application PC/TUS9200731  
; GENERAL INFORMATION:  
; APPLICANT: Kawabata, Shunichiro  
; APPLICANT: Davie, Earl W.  
; TITLE OF INVENTION: MICROSOMAL ENDOPEPTIDASE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: US  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/00731  
; FILING DATE: 19920128  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/646,997  
; FILING DATE: 28-JAN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 990008.415PC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206)622-4900  
; TELEFAX: (206)682-6031  
; TELEX: 3723836  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 709 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
PCT-US92-00731-13

Query Match 81.2%; Score 2979; DB 5; Length 709;  
Best Local Similarity 86.7%; Pred. No. 4.6e-264;  
Matches 568; Conservative 29; Mismatches 56; Indels 2; Gaps 1;

Qy 2 IARCLAVRSRLRRYGGSRILLRMTLGRVMSPLQAMSSYTVAGRNVLRWDLSPQIKTRT 61  
Db 54 IARCSAARGHLRRYGGSRILLRMTLGRVMSPLQAMSSYTAAGRNVLRWDLSPQIKTRT 113  
Qy 62 BELIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTMLDPHQHVSSEKVRAA 121  
Db 114 BELIAQTKQYDVGMLGIDKVTYENCLOALADVEVYIIVERTMLDPHQHVSSTREVRAA 173  
Qy 122 STEADKRLSRDIEMSMRGDIFFERIVHLOETCDLKGKTPKPEARRYLEKSIRKNGHLHP 181  
Db 174 STEADKRLSRDIEMSMRGDIFFERIVHLOETCDLKGKTPKPEARRYLEKSIRKNGHLHP 233  
Qy 182 EQVQNEIKSMKRMSELICIDFNKNLNEEDTFLVFSKAELGALPDDFIDSLKTDKDDYKI 241  
Db 234 XEVQNXIKSMKRMSELICIDFNKNLNEEDTFLVFSKAELGALPDDFIDSLKTDKDDYKI 293  
Qy 242 TLKYPHYFPVMMKKCCIPETRRRMEAFNTRCKEENTILQQLLPLRTKXAKLLGYSHAD 301  
Db 294 TLKYPHYFPVMMKKCCIPETRRRMEAFNTRCKEENTILQQLLPLRTKXAKLLGYSHAD 353  
Qy 302 FVLEMTAKSTSRVTAFLDLSQKLPLGEAREFELNKKCKDRGFEDYDGINAWDL 361  
Db 354 FVLEMTAKSTSRVTAFLDLSQKLPLGEAREFELNKKCKDRGFEDYDGINAWDL 413  
Qy 362 YYMTQTEELKYSIDQEFLEKYEFPFIEV--VTTEGLLNTYQELLGLSFQMTDAHVNMKSVT 419  
Db 414 HYMTQTEELKYSIDQEFLEKYEFPFIEV--VTTEGLLNTYQELLGLSFQMTDAHVNMKSVT 473  
Qy 420 LYTVKDKATGVLGQFYLDLYPREGKYNHAACTFGLQPCLLPDGSRMAYVYNFSQP 479  
Db 474 LYTVKDKATGVLGQFYLDLYPREGKYNHAACTFGLQPCLLPDGSRMAYVYNFSQP 533  
Qy 480 VAGPSLLRHDEVRTYFHEFGHVMHQICAOQDFARFSGTNTVETDFEVPVPSOMLENWYNDV 539  
Db 534 VAGPSLLRHDEVRTYFHEFGHVMHQICAOQDFARFSGTNTVETDFEVPVPSOMLENWYNDV 593  
Qy 540 DSLRRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASE 599  
Db 594 DSLRRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKVQDSLHTNTSLDAASE 653  
Qy 600 YAKYCEITLGAAPGTNNPATFGLAGGYGQYGYLWSEVFSMDMYSCFKKE 654  
Db 654 YARYCTDILGAAPGTNNPATFGLAGGYGQYGYLWSEVFSMDMYSCFKKE 708

RESULT 2  
US-07-766-351-5  
; Sequence 5, Application US/07766351  
; Patent No. 5292852  
; GENERAL INFORMATION:  
; APPLICANT: Sinha, Sukanto  
; APPLICANT: Seubert, Peter A.  
; APPLICANT: Dovey, Harry F.  
; APPLICANT: McConlogue, Lisa C.  
; APPLICANT: Little, Sheila P.  
; APPLICANT: Johnstone, Edward M.  
; TITLE OF INVENTION: Amyloidin Protease and Uses Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Athena Neurosciences, Inc.  
; STREET: 800F Gateway Blvd.  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07766.351  
; FILING DATE:

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy, Lisabeth Feix  
; REGISTRATION NUMBER: 31547  
; REFERENCE/DOCKET NUMBER: 17796-002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 877-0900  
; TELEFAX: (415) 877-8370  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 689 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; US-07-766-351-5  
  
Query Match 63.5%; Score 2331; DB 1; Length 689;  
Best Local Similarity 64.9%; Pred. No. 9.5e-205;  
Matches 426; Conservative 98; Mismatches 132; Indels 0; Gaps 0;  
  
Qy 46 NVLRWDLSPQIKTRTBEILIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTM 105  
Db 22 NDLRWDLSPQIKTRTBEILIVQKQYDVGMLGIEBVTYENCLOALADVEVYIIVERTM 81  
Qy 106 LDFFQHVSSDKRAAASAEADKRLSRDIEMSMRGDIFFERIVHLOETCDLKGKTPKPEARR 165  
Db 82 LDFFQHVSSDKRAAASAEADKRLSRDIEMSMRGDIFFERIVHLOETCDLKGKTPKPEARR 141  
Qy 166 LEKSIKMKRNGHLHPQVQNEIKSMKRMSELICIDFNKNLNEEDTFLVFSKAELGALPD 225  
Db 142 LERLIKLRNGHLHPQVQNEIKSMKRMSELICIDFNKNLNEEDTFLVFSKAELGALPD 201  
Qy 226 DFIDSLEKTDGKYLITLKYPHYPMKKCCIPETRRRMEAFNTRCKEENTILQQLL 285  
Db 202 DFLSLEKMDGKYLITLKYPHYPMKKCCIPETRRRMEAFNTRCKEENTILQQLL 261  
Qy 286 LRTVAKLLGYSHADFVLEMTAKSTSRVTAFLDLSQKLPLGEAREFELNKKCK 345  
Db 262 LRAQSRLLGLFHTADYVLENNMAKTSQTATFDELAKLPLGEQERAVILELKRAC 321  
Qy 346 KDRGFYDGNKINAWDLYYMTQTEELKYSIDQEFLEKYEFPFIEVTEGLLNTYQELLGLSF 405  
Db 322 ERRGLPFDGRIAWDMRYMNMQVEETRYCVDQNLLEKYEFPVQVWTHGLGIYQELLGLAF 381  
Qy 406 EQMTDAHVNMKSVTLTYVTKDKATGVLGQFYLDLYPREGKYNHAACTFGLQPCLLPDGSR 465  
Db 382 HHEGASAWHEDVRLYTARDAASGEVVGKFLDLYPREGKYGHAACTFGLQPCLLPDGSR 441  
Qy 466 MNAVALVYNFSQPVAGPSLLRHDEVRTYFHEFGHVMHQICAOQDFARFSGTNTVETDFV 525  
Db 442 QIATAAMVANFTKPTADAPSLQHDDEVETYPHEFGHVMHQLCSQAEAFMFGTHVERDFV 501  
Qy 526 EVPSOMLENWYNDVDSLRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKV 585  
Db 502 EAPSOLENWYNDVDSLRLSKHYKDGSPIDDLLEKLVASRLVNTGLTLRQIVLSKV 561  
Qy 586 QSLHTNTSLDAASEYAKYCEITLGAAPGTNNPATFGLAGGYGQYGYLWSEVFSMD 645  
Db 562 QALHTQTDADPAEYARLCQELLGPVATPGTNNPATFGLAGGYGQYGYLWSEVFSMD 621  
Qy 646 MFYSCFKKEGIMNPEVGMKYNLILKPGSGDLGMDMLHNLKREPNOKAFMLSRGL 701  
Db 622 MFHTFKEGVLNLSKVGMDYRSCILRPGSGEDASAMLRREFLGRDPKQDAFLLSKGL 677

RESULT 3  
US-08-059-032-5  
; Sequence 5, Application US/08059032  
; Patent No. 5424205  
; GENERAL INFORMATION:

Qy	46	NVLRWLDSPEQIKTRTEELIVOTKQYDVGMLGIEBVENTYENCQALADAVEKYIIVERTM	105
Db	22	NDLRWLDSAQIEERTRELIEQTKRYVDQVGTQEFEDSVESYESTKALADVEYTYVQVNI	81
Qy	106	LDPFQHVSSDKVEVRAASTEADKKLSRFDTEMSMRGDIFERIYVHQZTCDLGKPKPEAR	165
Db	82	LDPFQHVSPSKDITASTEADKKLSFEDVEMSMREDVYQRIYVWLQVKQKDSLRPEAARY	141
Qy	166	LEKSIKMGKNGHLHLPQVONEITKSMKKRMSLCLIDFNKLNLEDFFLVFSKAEALGALPD	225
Db	142	LERLIKLRNGRNGHLHLPQETQENIKRIKKLLSLCLIDFNKLNLEDFTTFLQELGGLPE	201
Qy	226	DFIDSLKTDDEKTKYTKLYPHYPYVMKKCIPETRRRMEMAFNTRCKEENTITLQOOLP	285
Db	202	DFLNSLEKMDGKLKVTLYKYPHYPYLLKCHVPETRRKVEEAFNCRCKEENCAILKELYT	261
Qy	286	LRTVKAKLGYSTHADFVLEMMNTAKTSRVTAFLLDLSQKRLPLGBAEREFILNKKKEC	345
Db	262	LRAQKSRLLGFTHADYVLEMMNAKTSQTVAFDLDELAQKLPLGQEERAVILELKRAEC	321
Qy	346	KDRGEYDVKINAWDYLYYNTQTEELKYSIDQEFLEKEYFPYEVWTEGLLNTYQELLGUSF	405
Db	322	ERRGLPFDGGRIRAWDMRYMYMNQVEETRYCVDQNLKKEYFPQVWTHGLLGIYQELLGLAF	381
Qy	406	EQMTDAHVKNSVTLITVKDKATGEVLGQFYLDLYIPREKGYNHAACFGLOPGLCLIPDGSR	465

```

1 FILING DATE: 19911004
2
3 CLASSIFICATION: 435
4
5 ATTORNEY/AGENT INFORMATION:
6
7 NAME: Murphy, Lisabeth Feix
8
9 REGISTRATION NUMBER: 31547
10
11 REFERENCE/DOCKET NUMBER: 17796-002
12
13 TELECOMMUNICATION INFORMATION:
14
15 TELEPHONE: (415) 877-0900
16
17 TELEFAX: (415) 877-8370
18
19 INFORMATION FOR SEQ ID NO: 5:
20
21 SEQUENCE CHARACTERISTICS:
22
23 LENGTH: 689 amino acids
24
25 TYPE: AMINO ACID
26
27 TOPOLOGY: linear
28
29 MOLECULE TYPE: protein
30
31 HYPOTHETICAL: NO
32
33 ANTI-SENSE: NO
34
35 PCT-US91-07290-5

```

Matches 426; Conservative 98; Mismatches 132; Indels 0; Gaps 0;

Qy 46 NVLRDLSPEQIKTRTEELIVQTKQYDDAVGMLGIEEYTYENCIALADVEVKYIVERTM 105  
 | | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : : | | | | | : : :  
 Db 22 NDLRDLSAQQIEERTRELIEQTKRYDQVGTFQEFEDVSYESITLAKALADVEVYTVQVNI 81

Qy	106	LDPPQHVSDDKEYEVAASTADKRLSRDIDEMSRGDIPIFETVHLQTCOLGKIKPEARRY	165
Db	82	LDPPHHVSPSSDKLTARTASTADKRLSDFEDVEMSRNEDVYQRIYVWLQEKVKQDSLRLPEAARY	141
Qy	166	LEKSIKMGKNGIHLPEOVQNETKSMKKRWSLCLIDFNKNLNEDDTFLVFSKAEALGALPD	225
Db	142	LEKLIKGRNGRHLPLPRETOENKRIKKLSLCLIDFNKNLNEDDTFLPFTLQELGLGPE	201
Qy	226	DFTDSLKTDDEKDKITLYKPHYPPYMKKCCIPETRRMEMAFNTRCKEENTIILOQLPL	285
Db	202	DFLNSLEKMGDKLVTKLYKPHYFPLKKCHVPETRRKYVEEAFNCRCKEENCALIKELVT	261
Qy	286	LRYKVAKLLGYSTHADPVLWEMNTAKTSRVTAPLDLDSOKLAPLPGBAEREFILNLKKKC	345
Db	262	LRAQKSLRLLGFTHADYVLEMNMAKTSQTATFLDELAKLPLGQERAVILELKRAC	321
Qy	346	KDRGFYDGRINAWDLVYVMTQTEELKYSIDQEFLEKYPFIEVVTGGLNTYOELLGLSF	405
Db	322	ERGLPFDGRIKAWDMRYVYMNQVEETRYCDQNLKKEYFPVQVTHGLLGIVOEELGLAF	381
Qy	406	EQMTDAHVWNKSYTLTVKDKATGEVLGYFDLYLPREGKYNHAACFGLQPGCLLPDGRS	465
Db	382	HHBEGASAWHEDVRLYTARDAASGEVVGKFLDLYPREGKYHAAACFGLQPGCLRQDGRS	441
Qy	466	MMAVAALVYVNFSPQVAGRSLLRBHDEVRIYFHEFGVHMHQICAQTOFARFSGTNVETDFV	525
Db	442	QIAIAAMVANFTKPTADAPSLQGHDEVETFYHEFGVHMQLCSQAFAFMFGTHVERDFV	501
Qy	526	EVPSQMLNNWVDVSLRRLSKYKDGSPITADDLLEKLVASRLVNTGLTLRQIVLSKYD	585
Db	502	EAPSQMLNNWVEQEPLLRMSRHYRTGSAVPRELLEKLTESQANTGLFSLRQIVLAKVD	561
Qy	586	QSHTNTSLDAASEYAKYCEIILGVAAATPGTNMPATFGHLAGGYDQYGYGLWSEVFSMD	645
Db	562	QALHTQTDADPAEYARLQCEILGVPTPGTNMPATFGHLAGGYDAQYGYGLWSEVFSMD	621
Qy	646	MFYSCKKEGIMPEVGMKYRNILIPGSGLDGMDMLHFLKREPQKQAFLLSRGL	701
Db	622	MFTTRKEQGVLSKNGMDYRSCILRPGGSEDASAMLRRLFLGRDPKQDAFLLSKGL	677
RESULT 5			
US-08-353-700-1			
; Sequence 1, Application US/08353700			
; Patent No. 5599919			
; GENERAL INFORMATION:			
; APPLICANT: YEN, TIMOTHY J.			
; APPLICANT: RATNER, JEROME B.			
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A			
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,			
; TITLE OF INVENTION: AND METHODS OF USE			
; NUMBER OF SEQUENCES: 4			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN			
; STREET: 1601 MARKET STREET, SUITE 720			
; CITY: PHILADELPHIA			
; STATE: PA			
; COUNTRY: USA			

```

; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; OS-08-353-700-1

Query Match 3.3%; Score 122; DB 1; Length 3248;
Best Local Similarity 19.8%; Pred. No. 0.22;
Matches 116; Conservative 91; Mismatches 238; Indels 142; Gaps

Qy 56 QIKTRTEELIVQKQVDVAGMLGIEETVYENCLOALAD--VEVYIYVERMLDPPQHVH 113
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 416 QMKARLTQELQAKNHH-----NVLOAELDKLTSVKQOLENNLEEFKQKLC 461

Qy 114 SDKEVRAASTEADKLRSRFDIEMSRMGDIFERIVHLOETCDLGKIKPARR--YLEKSIK 171
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 462 RASQATQASQIKENEURRMEEMKKNLTKS--HSEQ-----KAREVCHLEAEIK 510

Qy 172 MGRNRGLHLPQVQNKIKSMKKRMSELCDFNKNLNEDDTFLVFSKAEALGALPDPDFDSL 231
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 511 NIQK-CLNQSNFAEMKAKNTSQETMLRDLQEKINQOENSLTEKLKLA-----VADL 563

Qy 232 EKTDDDKYKITLKYHYFPMVKKCCIPETRRMEAFNT-----RCKEENTLIL-- 280
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 564 EKQDCSQDLLKREHHIQPLND-KLSKTEKESKALLSALELKKKEYBELKEETLFSCW 622

Qy 281 ----QQLPLRTKVKALLGYTHADFVLEMTAKSTSRVTAFLDOLQSKLPLGEARE 335
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 623 KSENEKLLTQMESEKENLQSKINHLCTUKTQOIKS-----HEYNERVTL-EMDRE 673

Qy 336 FTLNLKKKECKDRGFYDGGKINAWDLIYYTMTQTEELKYS-----IDQEFLEKFFPIE 387
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 674 --NL-----SVEIRNLHVLDSSKSVETQKLAYMELQQAEFSDQKHQKEINMC 722

Qy 388 VVTEGLLNTYQEL---LGLSFPQMTDAHVMKNSVTLYTVKDKATGEVLQGFYLDLYPREG 444
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 723 LKTSQLTQGVDELHKLQLLSNEIMD-----KDRC-----YQDL----- 756

Qy 445 KYNHAACFGLOPCLLPDGSRMVAALVVFNSQPVAGRPSSLRHDEVRTYFHEFGVHM 504
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 757 ----HAYESLURLLKSQD-----ASLVNEDH-----QRSLAFDQOPAMHHSFANIIG 802

Qy 505 QICAQTDFAFSGTNNVETDFVEVPSPOMLENVWVDLSLR-----RLSKHYKDKGS 553
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 803 EQGSMF--SERSECRLEADQSPKNSAILON---RVDLSFEFSLESQKMNSDLQCKOCEELV 857

Qy 554 PIADDLLEKLIVASRLVNTGLL--TLRQIVLSKVQDOSLHTNTSLDAAS 598
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 858 QIKGETEENLMKRAEOMHQSFAETSORISKLOEDTSAHONVVAETLS 904
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

RESULT      6
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Dann, Dorfman, Herrrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release 1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,700  
FILING DATE: 09-DEC-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: REED, JANET E.  
REGISTRATION NUMBER: 36,252  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 563-4100



```
/ STATE: PA
/ COUNTRY: USA
/ ZIP: 19103-2307
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/16216
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/353,700
/ FILING DATE: 09-DEC-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Reed, Janet E.
/ REGISTRATION NUMBER: 36,252
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 563-4100
/ TELEFAX: (215) 563-4044
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3248 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ PCT-US95-16216-1

Query Match 3.3%; Score 122; DB 5; Length 3248;
Best Local Similarity 19.8%; Pred. No. 0.22;
Matches 116; Conservative 91; Mismatches 238; Indels 142; Gaps 25;

QY 56 QIKTRTEELIVQTKQVYDAVGMLGIEBTVYENGLOALAD--VEVKYIVERTMLDFFQPHVS 113
Db 416 QMKARLTQELQQAQNMH-----NVLAELDKLTSVKQOLENNLEEFQKQKLC 461
QY 114 SKREVAARASTADKRLSRFDIEMSRGDIPIRIVHLQETCDLQKIKPEARR--YLEKSIX 171
Db 462 RAEQAFQASQIKENELRRSRMEEMKENLLKS--HSQ-----KAREVCHLEAELK 510
QY 172 MGRANGLHLPEQVONEIKSMKRMSELICDFNKNLNEDDTFLVFSKAEALGALPDDFDSL 231
Db 511 NIKQ-CLNQSQNFAEEMKAKNTSQETMLRDLQEKINQENSLTLEKILKA-----VADL 563
QY 232 EKTDDDKYKITLKYPHYPPVMKKCCIPETRRRMFAFNT-----RCKEENTIL-- 280
Db 564 EKORDCSQDLKKREHIEQIND-KISKTEKESKALLSLELAKKYEELKEKTLFSCW 622
QY 281 -----QQLPLRTKVAKLGYSTHADVLVEMTAKTSRVTAFDLDSLQSLKPLGEARE 335
Db 623 KSENEKLLTQWSEKENLQSKINHLETCLTKQIKS-----HEYNERNRTL-EMDRE 673
QY 336 FILNLKKKECKRGFEYDGINADWLIIYYTQTEELKYS-----IDQEFLEKEYPIE 387
Db 674 ---NL-----SVETRNILHNVLDSKSVETQKLAYMELQOAKFSDQKHQKEIENMC 722
QY 388 VVTEGLLNTYOEL--LGLSFEQMTDAHVNKSVTLYTVKDKATGVLGQFYLDLYPREG 444
Db 723 LKTSQLTQGVEDLEHLKQLLSNEIMD-----KDR-----YQDL----- 756
QY 445 KYNHAACFGQLPGCLLDGSRMMMAALVYVFSQPVAGRPSLLRLHRDVRVYTFHEFGVMH 504
Db 757 ---HAEYESRLDLKSKD-----ASLVNEDH---QKSLAFDQOPAMHHSFANIIG 802
QY 505 QICAQTDFAFSGTNVETDVEVPSQMLENNWVDVSLR-----RLSKHYKDG 553
Db 803 EQGSMP--SERSECRLEADQSPKNSAILQN---RVDSLEFSLESQKQMSDLQKQCELV 857

QY 554 PIADDLLEKIVASRLVNTGLL--TLRQIVLSKVDQSLTFTNTSLDAAS 598
Db 858 QIKGEIEENLMAEQMHQSFAETSQRISKLQEDTSAHQNVVAETLS 904

RESULT 7
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-576-1

Query Match 3.3%; Score 120; DB 2; Length 1388;
Best Local Similarity 19.8%; Pred. No. 0.084;
Matches 79; Conservative 83; Mismatches 158; Indels 78; Gaps 17;

QY 53 SPQIKTRTEELIVQTKQVYDAVGML-GIEE-VYENCLOALADVEVKYIVERTMLDPPQ 110
Db 629 SERRDRTHGSEII-----NDLQGRISGLEEDVKNKILLAKVELEKRLQER----- 675
QY 111 HVSSDKVEVRAASTADKRLSRFDIEMSRGDIPIRIVHLQETCDLQKIKPEARYLEK-- 168
Db 676 -----FTDLEKEKNNEIDMTYQKVIQOOSLEQET-----EHKATKARLADKNK 720
QY 169 ---SIKMGKRNGLHLPEQVONEIKSMKRMSELICDFNKNLNEDDTFLVFSKAEALGALPD 225
```

Db 721 IYSEIEAKSEAMKEMKKLSEBTLKQKVENLLEAEKRCSTLDCDLKOSQOKI-----N 776  
Qy 226 DFIDSLEKTDYKIKYIPYFVPMKCCIPETRRMEMAFNT-----RCKEENTII 279  
Db 777 ELLKQKQVNLNEDVRLNFKIEQ--ETQKRCGLTQNDLKMOTQVNTLKMSEKQLKQENHNL 834  
Qy 280 LQQLPLRTKAKLLGYSTHAD-----FVLENTAKST---SRVTAFLDDLKSLKPL 329  
Db 835 LEMKMSLEKQNAELRKERQADGOMKELQDLQLEAEQYFSTLYKTQVRELKBECEKTKLC 894  
Qy 330 GEABREILNKKKECKDRGFYDQKINAWDLYYMTQTELYKSIDQEFLEKVFPEV 389  
Db 895 KE-----LOOKKELQD-----ERSLAAQLEITUTKADSEOLARSIAEE---QISDLE-- 940  
Qy 390 TEGSLNTYQ--ELGLSFEQMTDAHVWNKSVTLTYVKD 425  
Db 941 KEKTMKELEIKEMMARHKQELTE-----KQATIASLEE 973  
RESULT 8  
US-07-977-434-12  
; Sequence 12, Application US/07977434  
; Patent No. 5466591  
; GENERAL INFORMATION:  
; APPLICANT: Gelfand, David H.  
; APPLICANT: Abramson, Richard D.  
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF  
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: New Jersey  
; ZIP: 07110-1199  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: 7  
; SOFTWARE: Wordperfect 2.1  
; CURRENT APPLICATION DATA: US/07977,434  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,490  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,466  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 590,213  
; FILING DATE: 28-SEP-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 523,394  
; FILING DATE: 15-MAY-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 143,441  
; FILING DATE: 12-JAN-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 063,509  
; FILING DATE: 17-JUN-1987  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 899,241  
; FILING DATE: 22-AUG-1986  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 746,121  
; FILING DATE: 15-AUG-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/US90/07641  
; FILING DATE: 21-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 585,471

; FILING DATE: 20-SEP-1990  
; PRIOR APPLICATION DATA: US 455,611  
; FILING DATE: 22-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 609,157  
; FILING DATE: 02-NOV-1990  
; PRIOR APPLICATION DATA: US 557,517  
; APPLICATION NUMBER: US 557,517  
; FILING DATE: 24-JUL-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Luann Cseri  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: Case No. 5466591 8753  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 814-2972  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 892 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-977-434-12  
  
Query Match 3.28; Score 116; DB 1; Length 892;  
Best Local Similarity 19.28; Pred. No. 0.095;  
Matches 118; Conservative 113; Mismatches 229; Indels 156; Gaps 30;  
  
Qy 85 YENGCLQALADVEVKYIVERTMLDFPQHVSDKEV-----RAASTEADKRLSRFDIEMS 137  
Db 31 HTNAVYGLTKMLIRFLKE-----HISICKDACVFLVDSKGGSKKRDILETYKANRP 82  
Qy 138 MRGD--IFERIVHLQETCD--LGKIKPEARRYLEKSIKMGKRNGLHLPQVQVQNEIKSMKRM 195  
Db 83 STPDLLLEQIPYVEELVDALG-----IKVLKIEGF-----EADDIIATLSKKF 125  
Qy 196 SELCIDFNKNLNEDDTFLVFSKAEALGALPDD--FIDSLEK--TD---DDKYITILKYPHY 248  
Db 126 ES---DFEK-----VNIITGDKOLLQVSDKVFVWRVVERGITDLVLYDRNKKVIEKGIY 176  
Qy 249 FPMVKKCCIPETRRMEMAFNTRCKEENTIILOQL-----LPLRTKVKAKLLGYSTHA 300  
Db 177 -----PE-----QFKDYLSDVGDQDINTPGYKIGKKTAVSLLKYNLSLE 216  
Qy 301 DFVLEMTAKSTSRVTAFLDDLKSLK---PLGEAERFILMLKKKECKDRGFYDQKIN 357  
Db 217 NVLKNINLL--TEKLRLLEDSDKEDLQKSIELVELIYDVPMDVEKDEIIYRGYNPDKLLK 274  
Qy 358 AWDLYYMTQTEELKYSIDQEFLEK---PTEVVTBGLLNT 396  
Db 275 VLKKEFFSSIIKEL--NLQEKLEKEYILVDNEDKLLKLAEELEKYEKTSIDTETTSLDPF 332  
Qy 397 YQELGLSFEQMTD-----AHVWNKSVTLTYVKDKATGEVLGQFYLDLYPREGKYNHA 449  
Db 333 EAKLVGISISTMEGKAYIPVSHFGAKNISKSLI--DKELQILQEKDYNIVGONLKFDYE 391  
Qy 450 --ACFGLOPG-----CLLPD-----GSRMMAVAALVNVFQPVAG 482  
Db 392 IFKSMGFSNPVPHFDITMIAAYLLNPDKFRNLEELSLKYLKYMISFDEL-VNENVPVLEF 450  
Qy 483 RP-SLLRHDEVITYPHEFGVHMQICAQTDFAFSGT-----NVETDFVEVPSOMLENN 535  
Db 451 NDFSVPPLERAVEYSCEDADVTRIFRLGRKIYENEMEKLFYEIMPLDVLSEMLNG 510  
Qy 536 VW-DVDSLRLRLSKHYKDGSPIDADDLEKL--VASRLVNTGLLT-LRQIVLSKVDOSLHTN 591  
Db 511 VYFDEEYKLSKKYQEK---MDGIKEKVEFETAGETFNLSSTQVAYILFEKINIAPYKK 567  
Qy 592 TSLDAASEYAKYCEI 607  
Db 568 TATGKFSTNAEVLEEL 583

```

RESULT 9
US-08-458-819-12
; Sequence 12, Application US/08458819
; Patent No. 5795762
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,819
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/977,434
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972

```

```

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-819-12

Query Match 3.28; Score 116; DB 1; Length 892;
Best local similarity 19.24; Pred. No. 0.095;
Matches 118; Conservative 113; Mismatches 229; Indels 156; Gaps 30;

QY 85 YENCLOALADVEVKYIVERTMLDPQHVSSDKV-----RAASTADKRLSRFDIEMS 137
DB 31 HTNAVYGLTKMLIKFLAE-----HISICKDACVFVLDKSGSKRKRKDKILEYKANRP 82
QY 138 MRGD-IFERIVHLOETCD-LGKIKPEARLYLEKSIKMGKRNGLHLPEQVQNEIKSMKRM 195
DB 83 STPDLLLEQIPYVEELVDALG-----IKVLKIEGF---EADDIIATLSKFK 125
QY 196 SELCIDENKLNEDDTFLVFSKALGALPDD--FIDSLEK--TD---DDKYKITLKYPHY 248
DB 126 ES----DFEK-----VNITGDKDLQLVSKVFWVRVERGITDLVLYDRNKVIEKYGIY 176
QY 249 FPMVKKCCIPETRRRMEAFNTRCKEENTIILOOL-----LPLRTKVKAKLLGYSTHA 300
DB 177 -----PE-----QFKDYLVLVDQIDNIPGVKGIGKKTAVSLLKKYNSLE 216
QY 301 DFVLEMTAKSTSRVTAFLDLSQKLK---PLGEAREFEILNKKKCKEKGDFEYDGKIN 357
DB 217 NVLKNINLL--TEKLRLLEDSEKEDLQKSIELVELIVDPMDVEKDEIYRGYNPKLLK 274
QY 358 ANDLIYMTQTEELKYSIDQFLKEY-----FPIEVVTEGLLNT 396
DB 275 VLKKEFESSIIEKEL--NLQEKLEKEYILVDNEDKLLKLAEEIEKYKTSIDTETSLDPF 332
QY 397 YQELLGLSFEQMTD-----AHVMNKSVTLYTVKDKATGEVLGQFYLDLPREGKYNHA 449
DB 333 EAKLVGISISTMECKAYIIPVSHEFGAKNISLSLI-DKFLKQILOEKDYNIVGQNLKFDYE 391
QY 450 --ACFGLQPG-----CLLPD-----GSRMMAVAALVYVNSQPVAG 482
DB 392 IFKSMGFSPNVPHPDITMAAYLLNPDEKRFNLEELSLKYLGYKMISFDEL-VNENVPFLG 450
QY 483 RP-SLLRHDEVRTYFHEFGVHMHQICATDFARESGT-----NVEDTFVEVPQOMLENW 535
DB 451 NDFSIVPLERAVEYSCEDADVTYRFRKLRKRIYENEMEKLFYEIEMPLIDVLSEMEUNG 510
QY 536 VW-DVDSLRLSLRKHKDGSPIDADDLLEKL--VASRLVNTGLLT-LRQIVLSKVDQSHTN 591
DB 511 VYFDEEVLKLSKYYQEK---MDGIKEKVFPIAGETFNLSSTQVAYLLFEKLNAPYKK 567
QY 592 TSLDAASEYAKYCSEI 607
DB 568 TATGKFSTNAEVLLEL 583

RESULT 10
PCT-US91-07035-12
; Sequence 12, Application PC/TUS9107035
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cetus Corporation
; STREET: 1400 Fifty-third Street
; CITY: Emeryville
; STATE: California
; ZIP: 94608

```





APPLICANT: Davis, Maria  
APPLICANT: Sha, Dan  
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
SUITE: Suite 4700  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/766,014  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below:  
APPLICATION NUMBER: 60/008,688  
FILING DATE: December 15, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Warburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 223/104  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 872 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-766-014-2

Query Match 3.1%; Score 115; DB 1; Length 872;  
Best Local Similarity 17.6%; Pred. No. 0.11;  
Matches 108; Conservative 115; Mismatches 195; Indels 196; Gaps 27;  
Qy 54 PEQKTRTELIVQKQYDVGMLGIEEVTYE-----NCIQALAD---VEVKYIV-ERT 104  
Db 80 PEEAAEQVDYL---KEIIDGNFKITLEGEYADDDIIGTISKLAEEKGMEVLVVTGDRD 135  
Qy 105 MLDPEQVSSDKVRAASTADKRLSRDIEMSRGDIFERI-VHLQETCDLKGKIPPEAR 163  
Db 136 AL--QVSKVQKIKSKGKITQMEFEDEKA-----ILERYGITPOQFIDLGLMGDKS 186  
Qy 164 RYLEKSIKMKRNGHL-----PEQVQN--EIKSMKRKMSLELC- 199  
Db 187 DNIQVNPNGEIKTAIKLLKDFGTIENLQNLSQLKGIENIENKLAISKRLATIKR 246  
Qy 200 -----IDFN-----KNLNEDDTFLVFSKAEALPDFTDLSLEKTDKDKYKTLKYPHYFP 250  
Db 247 DIPIEDPEEYKVKFNEEKLLELFNKEFTSL-----IDNIKESS----- 288  
Qy 251 VMKKCCIPETRRRMEAMTRCKEENTIILOQLPLRTPKVAKLLGYSTHADFLVEMNTAK 310  
Db 289 -----TEIVDNHVKVSKVVDIKELVTLQDNRNIAFYPLIYEGEIK-KIAF 334  
Qy 311 STSRVTAFLDDL-SQKPLGEGAE-----REFINLAKK--ECKDRGFEDGKINA 358  
Db 335 SFGKDTVIDVFQEDLKEIFEKEDFEFTTHEIKDFLVRLSYKGIECKSK--YIDTAVMA 392  
Qy 359 WDLYYMTQTTELKYSIDQEFLEKFFPIEVVVT-EGLLNTYOELLGLS--FQGMTDAHVMN 415

Db 393 -----YLLNPSESNYDLDR-VLKYLKVDVPSYEGIFGKGRKKKIEIDENILADYICS 446  
Qy 416 KSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAACFGLOPCLLPDGSRRMAVAALVWN 475  
Db 447 RCYVFLDKLKEK-----LMNFIEMDMKKLLLE 473  
Qy 476 FSQPVAGRPSLLRHDEVRTYFHFEGHVMHQICAOQDFARFSGTNTVETDFVEVPSQMLEN- 534  
Db 474 IEMPLV---EVLKSMEV-----SGFTLDKEVLKELSKQIDDR 507  
Qy 535 -----WVDVDSLRRLS-----KHFKDGSPIADDLLEKLVASRLV 569  
Db 508 IGEILDKIYKAGYQFNWNSPKQLSEFLFEKLNLPVKKTKTGSTYDSEVLEQLVPYNDI 567  
Qy 570 NTGLLTQLRQIVLSK 583  
Db 568 VSDIIEYRQLTKLK 581  
RESULT 14  
US-08-245-511-47  
Sequence 47, Application US/08245511  
Patent No. 5928900  
GENERAL INFORMATION:  
APPLICANT: Measure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 01-SEP-1994  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 643 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-245-511-47

Query Match 3.1%; Score 113.5; DB 2; Length 643;  
Best Local Similarity 19.4%; Pred. No. 0.095;  
Matches 86; Conservative 57; Mismatches 152; Indels 149; Gaps 16;  
Qy 296 YSTHADFLVEMNTAK-STSRVTAFLDDLSQLKPLGAEAREFILLNKKKCKECKRGFEYDG 354

Db 23 YETDPNLNLYLTAKAANTITSNVVD-----GLLENDRYGNFVPSMAEDWSYKDG 74  
QY 355 KINAW-----DLYYMTQTEE-----LKYSIDQEFLEKYEPIEVVTEGLLNTYQ 398  
Db 75 LYVTYIRKDAKWYTSGEGEYAAVKAQDFVTGLKYAADKK--SDALYPVOESIKGL----- 128  
QY 399 ELLGLSFEQMTDAHVNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAAACFGL----- 454  
Db 129 -----DAYVKEIKDFSQVGIKALDEQTVQYTLN--KPEFWNSKTTMGVLAPV 175  
QY 455 -----QPCLLPDGSRRMVAALVYNFVSQVAGRPSSLRHRHDEVRTYFHE 498  
Db 176 NEEFLNSKGGDFAKATDPSSLLYNGPYLL--KSIYTKSSVEFAKNPNYWDKD----- 225  
QY 499 FGHVMHQICAQDFARFSGTNETDFVEVPSQMLENNWVDVSLRLSKHYKDGSP1ADD 558  
Db 226 -----NVHIDKVKLS-----FWDGQDTSKPAENFKDGSILTAAR 258  
QY 559 L-----LEKLVASRLVNTGLTLRLQIVLSKVDQSLHTNTSLDAASEYAK----- 602  
Db 259 LYPTSASFAELEKSMKDNIVYTQDSITVLTGTNIDRQSYKYTSKTSDEQKASTKKALLN 318  
QY 603 -----YCSILGVAATP-----GTNMPATFGLAGGYDGY-----YGY 636  
Db 319 KDFRQIAFGFDRATAYASQLNGTGASKILRLNLFVPTFVQADGKNFGDMVKEKLVTYGD 378  
QY 637 LMSEVFSMDMFYSCFKKEGIMNPE 660  
Db 379 EWKDVNLAD-----SQDGLYNPE 396

RESULT 15

US-08-600-993A-47

; Sequence 47, Application US/08600993A  
; Patent No. 5981229  
; GENERAL INFORMATION:  
; APPLICANT: Masure, H Robert  
; APPLICANT: Pearce, Barbara J  
; APPLICANT: Tuomanen, Elaine  
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/600,993A  
; FILING DATE: 1-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/245,511  
; FILING DATE: 01-SEP-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-069 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 643 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-600-993A-47

Query Match 3.1%; Score 113.5; DB 2; Length 643;  
Best local Similarity 19.4%; Pred. No. 0.095;  
Matches 86; Conservative 57; Mismatches 152; Indels 149; Gaps 16;  
QY 296 YSTHADFLVEMTAK--STSRVTAFLDLDLQSLKPLCEAEREFLNLKKKECKDRGEYDG 354  
Db 23 YETDPNLNLYLTAKAANTITSNVVD-----GLLENDRYGNFVPSMAEDWSYKDG 74  
QY 355 KINAW-----DLYYMTQTEE-----LKYSIDQEFLEKYEPIEVVTEGLLNTYQ 398  
Db 75 LYVTYIRKDAKWYTSGEGEYAAVKAQDFVTGLKYAADKK--SDALYPVOESIKGL----- 128  
QY 399 ELLGLSFEQMTDAHVNKSVTLTYTKDKATGEVLGQFYLDLYPREGKYNHAAACFGL----- 454  
Db 129 -----DAYVKEIKDFSQVGIKALDEQTVQYTLN--KPEFWNSKTTMGVLAPV 175  
QY 455 -----QPCLLPDGSRRMVAALVYNFVSQVAGRPSSLRHRHDEVRTYFHE 498  
Db 176 NEEFLNSKGGDFAKATDPSSLLYNGPYLL--KSIYTKSSVEFAKNPNYWDKD----- 225  
QY 499 FGHVMHQICAQDFARFSGTNETDFVEVPSQMLENNWVDVSLRLSKHYKDGSP1ADD 558  
Db 226 -----NVHIDKVKLS-----FWDGQDTSKPAENFKDGSILTAAR 258  
QY 559 L-----LEKLVASRLVNTGLTLRLQIVLSKVDQSLHTNTSLDAASEYAK----- 602  
Db 259 LYPTSASFAELEKSMKDNIVYTQDSITVLTGTNIDRQSYKYTSKTSDEQKASTKKALLN 318  
QY 603 -----YCSILGVAATP-----GTNMPATFGLAGGYDGY-----YGY 636  
Db 319 KDFRQIAFGFDRATAYASQLNGTGASKILRLNLFVPTFVQADGKNFGDMVKEKLVTYGD 378  
QY 637 LMSEVFSMDMFYSCFKKEGIMNPE 660  
Db 379 EWKDVNLAD-----SQDGLYNPE 396

Search completed: January 14, 2002, 18:56:33  
Job time: 792 sec

